

(51) International Patent Classification : C12N 15/16, C07K 14/72	A2	(11) International Publication Number: 20 April 2000 (20.04.00)
	(43) International Publication Date:	

(23) International Application Number: PCT/US99/24065	(72) Inventors, and (75) Inventor/applicants (for US only): BEHAN, Dominic, P. (GB/US); 11472 Roberto Court, San Diego, CA 92131 (US); LEHMANN-BRUNSMMA, Karin (DE/US); 1265 Paines Lane, San Diego, CA 92139 (US); CHAMBERS, Dextra, T. (GB/US); 347 Longdon Lane, Solina Beach, CA 92150 (US); CHEN, Ruoping (CN/US); 5296 Timber Branch Way, San Diego, CA 92130 (US); DANG, Hong, T. (US/US); 5532 Oak Park Drive, San Diego, CA 92105 (US); GORE, Martin (GB/US); 6668 Eutaw Avenue, San Diego, CA 92120 (US); LIAW, Chin, W. (US/US); 7668 Santa Place, San Diego, CA 92129 (US); LIN, Fahn-I (US); 8291-7 Gold Coast Drive, San Diego, CA 92126 (US); LOWITZ, Kevin (US/US); Apartment C, 8031 Camino de Pizzo, San Diego, CA 92108 (US); WHITE, Carol (US/US); 4260 Cleveland Avenue, San Diego, CA 92105 (US).
(22) International Filing Date: 13 October 1999 (13.10.99)	(74) Agents: MILLER, Suzanne, E. et al.; Woodcock Washburn Kunz Maciewicz & North LLP, 46th floor, One Liberty Place, Philadelphia, PA 19103 (US).
(30) Priority Data: 09/170,496 60/108,029 60/109,215 60/109,260 60/109,261 60/120,416 60/120,417 60/123,944 60/123,945 60/123,948 60/123,946 60/123,949 60/123,951 60/123,952 60/123,956 60/123,957 60/123,958 60/123,959 60/123,960 60/123,961 60/123,962 60/123,963 60/123,964 60/123,965 60/123,966 60/123,967 60/123,968 60/123,969 60/123,970 60/123,971 60/123,972 60/123,973 60/123,974 60/123,975 60/123,976 60/123,977 60/123,978 60/123,979 60/123,980 60/123,981 60/123,982 60/123,983 60/123,984 60/123,985 60/123,986 60/123,987 60/123,988 60/123,989 60/123,990 60/123,991 60/123,992 60/123,993 60/123,994 60/123,995 60/123,996 60/123,997 60/123,998 60/123,999 60/124,000 60/124,001 60/124,002 60/124,003 60/124,004 60/124,005 60/124,006 60/124,007 60/124,008 60/124,009 60/124,010 60/124,011 60/124,012 60/124,013 60/124,014 60/124,015 60/124,016 60/124,017 60/124,018 60/124,019 60/124,020 60/124,021 60/124,022 60/124,023 60/124,024 60/124,025 60/124,026 60/124,027 60/124,028 60/124,029 60/124,030 60/124,031 60/124,032 60/124,033 60/124,034 60/124,035 60/124,036 60/124,037 60/124,038 60/124,039 60/124,040 60/124,041 60/124,042 60/124,043 60/124,044 60/124,045 60/124,046 60/124,047 60/124,048 60/124,049 60/124,050 60/124,051 60/124,052 60/124,053 60/124,054 60/124,055 60/124,056 60/124,057 60/124,058 60/124,059 60/124,060 60/124,061 60/124,062 60/124,063 60/124,064 60/124,065 60/124,066 60/124,067 60/124,068 60/124,069 60/124,070 60/124,071 60/124,072 60/124,073 60/124,074 60/124,075 60/124,076 60/124,077 60/124,078 60/124,079 60/124,080 60/124,081 60/124,082 60/124,083 60/124,084 60/124,085 60/124,086 60/124,087 60/124,088 60/124,089 60/124,090 60/124,091 60/124,092 60/124,093 60/124,094 60/124,095 60/124,096 60/124,097 60/124,098 60/124,099 60/124,100 60/124,101 60/124,102 60/124,103 60/124,104 60/124,105 60/124,106 60/124,107 60/124,108 60/124,109 60/124,110 60/124,111 60/124,112 60/124,113 60/124,114 60/124,115 60/124,116 60/124,117 60/124,118 60/124,119 60/124,120 60/124,121 60/124,122 60/124,123 60/124,124 60/124,125 60/124,126 60/124,127 60/124,128 60/124,129 60/124,130 60/124,131 60/124,132 60/124,133 60/124,134 60/124,135 60/124,136 60/124,137 60/124,138 60/124,139 60/124,140 60/124,141 60/124,142 60/124,143 60/124,144 60/124,145 60/124,146 60/124,147 60/124,148 60/124,149 60/124,150 60/124,151 60/124,152 60/124,153 60/124,154 60/124,155 60/124,156 60/124,157 60/124,158 60/124,159 60/124,160 60/124,161 60/124,162 60/124,163 60/124,164 60/124,165 60/124,166 60/124,167 60/124,168 60/124,169 60/124,170 60/124,171 60/124,172 60/124,173 60/124,174 60/124,175 60/124,176 60/124,177 60/124,178 60/124,179 60/124,180 60/124,181 60/124,182 60/124,183 60/124,184 60/124,185 60/124,186 60/124,187 60/124,188 60/124,189 60/124,190 60/124,191 60/124,192 60/124,193 60/124,194 60/124,195 60/124,196 60/124,197 60/124,198 60/124,199 60/124,200 60/124,201 60/124,202 60/124,203 60/124,204 60/124,205 60/124,206 60/124,207 60/124,208 60/124,209 60/124,210 60/124,211 60/124,212 60/124,213 60/124,214 60/124,215 60/124,216 60/124,217 60/124,218 60/124,219 60/124,220 60/124,221 60/124,222 60/124,223 60/124,224 60/124,225 60/124,226 60/124,227 60/124,228 60/124,229 60/124,230 60/124,231 60/124,232 60/124,233 60/124,234 60/124,235 60/124,236 60/124,237 60/124,238 60/124,239 60/124,240 60/124,241 60/124,242 60/124,243 60/124,244 60/124,245 60/124,246 60/124,247 60/124,248 60/124,249 60/124,250 60/124,251 60/124,252 60/124,253 60/124,254 60/124,255 60/124,256 60/124,257 60/124,258 60/124,259 60/124,260 60/124,261 60/124,262 60/124,263 60/124,264 60/124,265 60/124,266 60/124,267 60/124,268 60/124,269 60/124,270 60/124,271 60/124,272 60/124,273 60/124,274 60/124,275 60/124,276 60/124,277 60/124,278 60/124,279 60/124,280 60/124,281 60/124,282 60/124,283 60/124,284 60/124,285 60/124,286 60/124,287 60/124,288 60/124,289 60/124,290 60/124,291 60/124,292 60/124,293 60/124,294 60/124,295 60/124,296 60/124,297 60/124,298 60/124,299 60/124,300 60/124,301 60/124,302 60/124,303 60/124,304 60/124,305 60/124,306 60/124,307 60/124,308 60/124,309 60/124,310 60/124,311 60/124,312 60/124,313 60/124,314 60/124,315 60/124,316 60/124,317 60/124,318 60/124,319 60/124,320 60/124,321 60/124,322 60/124,323 60/124,324 60/124,325 60/124,326 60/124,327 60/124,328 60/124,329 60/124,330 60/124,331 60/124,332 60/124,333 60/124,334 60/124,335 60/124,336 60/124,337 60/124,338 60/124,339 60/124,340 60/124,341 60/124,342 60/124,343 60/124,344 60/124,345 60/124,346 60/124,347 60/124,348 60/124,349 60/124,350 60/124,351 60/124,352 60/124,353 60/124,354 60/124,355 60/124,356 60/124,357 60/124,358 60/124,359 60/124,360 60/124,361 60/124,362 60/124,363 60/124,364 60/124,365 60/124,366 60/124,367 60/124,368 60/124,369 60/124,370 60/124,371 60/124,372 60/124,373 60/124,374 60/124,375 60/124,376 60/124,377 60/124,378 60/124,379 60/124,380 60/124,381 60/124,382 60/124,383 60/124,384 60/124,385 60/124,386 60/124,387 60/124,388 60/124,389 60/124,390 60/124,391 60/124,392 60/124,393 60/124,394 60/124,395 60/124,396 60/124,397 60/124,398 60/124,399 60/124,400 60/124,401 60/124,402 60/124,403 60/124,404 60/124,405 60/124,406 60/124,407 60/124,408 60/124,409 60/124,410 60/124,411 60/124,412 60/124,413 60/124,414 60/124,415 60/124,416 60/124,417 60/124,418 60/124,419 60/124,420 60/124,421 60/124,422 60/124,423 60/124,424 60/124,425 60/124,426 60/124,427 60/124,428 60/124,429 60/124,430 60/124,431 60/124,432 60/124,433 60/124,434 60/124,435 60/124,436 60/124,437 60/124,438 60/124,439 60/124,440 60/124,441 60/124,442 60/124,443 60/124,444 60/124,445 60/124,446 60/124,447 60/124,448 60/124,449 60/124,450 60/124,451 60/124,452 60/124,453 60/124,454 60/124,455 60/124,456 60/124,457 60/124,458 60/124,459 60/124,460 60/124,461 60/124,462 60/124,463 60/124,464 60/124,465 60/124,466 60/124,467 60/124,468 60/124,469 60/124,470 60/124,471 60/124,472 60/124,473 60/124,474 60/124,475 60/124,476 60/124,477 60/124,478 60/124,479 60/124,480 60/124,481 60/124,482 60/124,483 60/124,484 60/124,485 60/124,486 60/124,487 60/124,488 60/124,489 60/124,490 60/124,491 60/124,492 60/124,493 60/124,494 60/124,495 60/124,496 60/124,497 60/124,498 60/124,499 60/124,500 60/124,501 60/124,502 60/124,503 60/124,504 60/124,505 60/124,506 60/124,507 60/124,508 60/124,509 60/124,510 60/124,511 60/124,512 60/124,513 60/124,514 60/124,515 60/124,516 60/124,517 60/124,518 60/124,519 60/124,520 60/124,521 60/124,522 60/124,523 60/124,524 60/124,525 60/124,526 60/124,527 60/124,528 60/124,529 60/124,530 60/124,531 60/124,532 60/124,533 60/124,534 60/124,535 60/124,536 60/124,537 60/124,538 60/124,539 60/124,540 60/124,541 60/124,542 60/124,543 60/124,544 60/124,545 60/124,546 60/124,547 60/124,548 60/124,549 60/124,550 60/124,551 60/124,552 60/124,553 60/124,554 60/124,555 60/124,556 60/124,557 60/124,558 60/124,559 60/124,560 60/124,561 60/124,562 60/124,563 60/124,564 60/124,565 60/124,566 60/124,567 60/124,568 60/124,569 60/124,570 60/124,571 60/124,572 60/124,573 60/124,574 60/124,575 60/124,576 60/124,577 60/124,578 60/124,579 60/124,580 60/124,581 60/124,582 60/124,583 60/124,584 60/124,585 60/124,586 60/124,587 60/124,588 60/124,589 60/124,590 60/124,591 60/124,592 60/124,593 60/124,594 60/124,595 60/124,596 60/124,597 60/124,598 60/124,599 60/124,600 60/124,601 60/124,602 60/124,603 60/124,604 60/124,605 60/124,606 60/124,607 60/124,608 60/124,609 60/124,610 60/124,611 60/124,612 60/124,613 60/124,614 60/124,615 60/124,616 60/124,617 60/124,618 60/124,619 60/124,620 60/124,621 60/124,622 60/124,623 60/124,624 60/124,625 60/124,626 60/124,627 60/124,628 60/124,629 60/124,630 60/124,631 60/124,632 60/124,633 60/124,634 60/124,635 60/124,636 60/124,637 60/124,638 60/124,639 60/124,640 60/124,641 60/124,642 60/124,643 60/124,644 60/124,645 60/124,646 60/124,647 60/124,648 60/124,649 60/124,650 60/124,651 60/124,652 60/124,653 60/124,654 60/124,655 60/124,656 60/124,657 60/124,658 60/124,659 60/124,660 60/124,661 60/124,662 60/124,663 60/124,664 60/124,665 60/124,666 60/124,667 60/124,668 60/124,669 60/124,670 60/124,671 60/124,672 60/124,673 60/124,674 60/124,675 60/124,676 60/124,677 60/124,678 60/124,679 60/124,680 60/124,681 60/124,682 60/124,683 60/124,684 60/124,685 60/124,686 60/124,687 60/124,688 60/124,689 60/124,690 60/124,691 60/124,692 60/124,693 60/124,694 60/124,695 60/124,696 60/124,697 60/124,698 60/124,699 60/124,700 60/124,701 60/124,702 60/124,703 60/124,704 60/124,705 60/124,706 60/124,707 60/124,708 60/124,709 60/124,710 60/124,711 60/124,712 60/124,713 60/124,714 60/124,715 60/124,716 60/124,717 60/124,718 60/124,719 60/124,7	

The invention disclosed in this patent document relates to transmembrane receptors, more particularly to a human G protein-coupled receptor for which the endogenous ligand is unknown ("orphan GPCR receptors"), and most particularly to mutated (non-endogenous) versions of the human GPCRs for evidence of constitutive activity.

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT

AL	Albania	BS	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	PT	Portugal	SK	Slovakia	SK	Slovenia
AN	Antigua	FR	France	LN	Liechtenstein	SN	Senegal
AV	Austria	OA	Oman	LV	Latvia	SZ	Switzerland
AU	Australia	GB	United Kingdom	MC	Monaco	TD	Togo
AZ	Azerbaijan	GE	Georgia	MD	Republic of Moldova	TC	Turkey
BA	Bosnia and Herzegovina	GH	Ghana	TJ	Tajikistan	TM	Turkmenistan
BB	Barbados	GR	Greece	MG	Madagascar	TR	Turkey
BE	Belgium	OM	Oman	MK	The former Yugoslav Republic of Macedonia	TT	Trinidad and Tobago
BF	Burkina Faso	OR	Uruguay	ML	Mali	UA	Ukraine
BG	Bulgaria	HU	Hungary	NN	Norway	UG	Uganda
BJ	Benin	IE	Ireland	MR	Mauritania	US	United States of America
BR	Brazil	IL	Israel	MW	Malawi	UZ	Uzbekistan
BY	Belarus	IS	Iceland	MX	Mexico	VN	Viet Nam
CA	Canada	IT	Italy	NI	Niger	VO	Viet Nam
CF	Central African Republic	JP	Japan	NL	Netherlands	VU	Vanuatu
CG	Congo	KE	Kenya	NZ	New Zealand	ZW	Zimbabwe
CH	Switzerland	KG	Kyrgyzstan				
CI	Cote d'Ivoire	KR	Democratic People's Republic of Korea				
CM	Cameroun	KZ	Kazakhstan				
CN	China	LA	Laos				
CU	Cuba	KR	Republic of Korea				
CV	Cape Verde	KZ	Kazakhstan				
CZ	Czech Republic	LC	Saint Lucia				
DE	Germany	LI	Liechtenstein				
DK	Denmark	LK	Sri Lanka				
EE	Estonia	LR	Liberia				

- 2 -

NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED HUMAN G PROTEIN-COUPLED RECEPTORS

This patent application is a continuation-in-part of, and claims priority from, U.S.

Serial Number 09/170,496, filed with the United States Patent and Trademark Office on

5 October 13, 1998. This application also claims the benefit of priority from the following provisional applications, all filed via U.S. Express Mail with the United States Patent and Trademark Office on the indicated dates: U.S. Provisional Number 60/110,060, filed November 27, 1998; U.S. Provisional Number 60/120,416, filed February 16, 1999; U.S. Provisional Number 60/121,852, filed February 26, 1999 claiming benefit of U.S.

10 Provisional Number 60/109,213, filed November 20, 1998; U.S. Provisional Number 60/123,944, filed March 12, 1999; U.S. Provisional Number 60/123,945, filed March 12, 1999; U.S. Provisional Number 60/123,948, filed March 12, 1999; U.S. Provisional Number 60/123,946, filed March 12, 1999; U.S. Provisional Number 60/123,949, filed March 12, 1999; U.S.

15 Provisional Number 60/152,524, filed September 3, 1999, claiming benefit of U.S. Provisional Number 60/151,114, filed August 27, 1999 and U.S. Provisional Number 60/108,029, filed November 12, 1998; U.S. Provisional Number 60/136,436, filed May 28, 1999; U.S. Provisional Number 60/136,439, filed May 28, 1999; U.S. Provisional Number 60/136,567, filed May 28, 1999; U.S. Provisional Number 60/137,127, filed May 28,

20 1999; U.S. Provisional Number 60/137,131, filed May 28, 1999; U.S. Provisional Number

60/141,448, filed June 29, 1999 claiming benefit of U.S. Provisional Number 60/136,437, filed May 28, 1999; U.S. Provisional Number 60/156,633, filed September 29, 1999; U.S. Provisional Number 60/156,555, filed September 29, 1999; U.S. Provisional Number 60/156,634, filed September 29, 1999; U.S. Provisional Number ____ (Arena

5 Pharmaceuticals, Inc. docket number: CHN10-1), filed September 29, 1999; U.S.

Provisional Number ____ (Arena Pharmaceuticals, Inc. docket number: RUP6-1), filed October 1, 1999; U.S. Provisional Number ____ (Arena Pharmaceuticals, Inc. docket number: RUP7-1), filed October 1, 1999; U.S. Provisional Number ____ (Arena

10 Pharmaceuticals, Inc. docket number: CHN6-1), filed October 1, 1999; U.S. Provisional Number ____ (Arena Pharmaceuticals, Inc. docket number: RUP5-1), filed October 1, 1999; and U.S. Provisional Number ____ (Arena Pharmaceuticals, Inc. docket number: CHN9-1), filed October 1, 1999. This application is also related to co-pending U.S. Serial Number ____ (Woodcock, Washburn, Kurtz, Makiewicz & Norris, LLP docket number AREN-

15 0050), filed on October 12, 1999 (via U.S. Express Mail) and U.S. Serial Number 09/364,425, filed on July 30, 1999, both incorporated herein by reference. This application also claims priority to U.S. Serial Number ____ (Woodcock, Washburn, Kurtz, Makiewicz & Norris, LLP docket number AREN-0054), filed on October 12, 1999 (via U.S. Express Mail), incorporated by reference herein in its entirety. Each of the foregoing applications are incorporated by reference herein in their entirety.

20

FIELD OF THE INVENTION

The invention disclosed in this patent document relates to transmembrane receptors, and more particularly to human G protein-coupled receptors, and specifically to

- 3 -

GPCRs that have been altered to establish or enhance constitutive activity of the receptor. Preferably, the altered GPCRs are used for the direct identification of candidate compounds as receptor agonists, inverse agonists or partial agonists having potential applicability as therapeutic agents.

BACKGROUND OF THE INVENTION

Although a number of receptor classes exist in humans, by far the most abundant and therapeutically relevant is represented by the G protein-coupled receptor (GPCR or GPCRs) class. It is estimated that there are some 100,000 genes within the human genome, and of these, approximately 2%, or 2,000 genes, are estimated to code for GPCRs. Receptors, including GPCRs, for which the endogenous ligand has been identified are referred to as "known" receptors, while receptors for which the endogenous ligand has not been identified are referred to as "orphan" receptors. GPCRs represent an important area for the development of pharmaceutical products: from approximately 20 of the 100 known GPCRs, 60% of all prescription pharmaceuticals have been developed.

GPCRs share a common structural motif. All these receptors have seven sequences of between 22 to 24 hydrophobic amino acids that form seven alpha helices, each of which spans the membrane (each span is identified by number, *i.e.*, transmembrane-1 (TM-1), transmembrane-2 (TM-2), etc.). The transmembrane helices are joined by strands of amino acids between transmembrane-2 and transmembrane-3, transmembrane-4 and transmembrane-5, and transmembrane-6 and transmembrane-7 on the exterior, or "extracellular" side, of the cell membrane (these are referred to as "extracellular" regions 1, 2 and 3 (EC-1, EC-2 and EC-3), respectively). The transmembrane helices are also joined by strands of amino acids between transmembrane-1 and transmembrane-2, transmembrane-3 and transmembrane-4, and

- 4 -

transmembrane-5 and transmembrane-6 on the interior, or "intracellular" side, of the cell membrane (these are referred to as "intracellular" regions 1, 2 and 3 (IC-1, IC-2 and IC-3), respectively). The "carboxy" ("C") terminus of the receptor lies in the intracellular space within the cell, and the "amino" ("N") terminus of the receptor lies in the extracellular space outside of the cell.

Generally, when an endogenous ligand binds with the receptor (often referred to as "activation" of the receptor), there is a change in the conformation of the intracellular region that allows for coupling between the intracellular region and an intracellular "G-protein." It has been reported that GPCRs are "promiscuous" with respect to G proteins, *i.e.*, that a GPCR can interact with more than one G protein. *See, Kenakin, T., 43 Life Sciences 1095 (1988).* Although other G proteins exist, currently, Gq, Gs, Gi, Gz and Go are G proteins that have been identified. Endogenous ligand-activated GPCR coupling with the G-protein begins a signaling cascade process (referred to as "signal transduction"). Under normal conditions, signal transduction ultimately results in cellular activation or cellular inhibition. It is thought that the IC-3 loop as well as the carboxy terminus of the receptor interact with the G protein.

Under physiological conditions, GPCRs exist in the cell membrane in equilibrium between two different conformations: an "inactive" state and an "active" state. A receptor in an inactive state is unable to link to the intracellular signaling transduction pathway to produce a biological response. Changing the receptor conformation to the active state allows linkage to the transduction pathway (via the G-protein) and produces a biological response.

A receptor may be stabilized in an active state by an endogenous ligand or a

compound such as a drug. Recent discoveries, including but not exclusively limited to modifications to the amino acid sequence of the receptor, provide means other than endogenous ligands or drugs to promote and stabilize the receptor in the active state conformation. These means effectively stabilize the receptor in an active state by simulating the effect of an endogenous ligand binding to the receptor. Stabilization by such ligand-independent means is termed "constitutive receptor activation."

SUMMARY OF THE INVENTION

Disclosed herein are non-endogenous versions of endogenous, human GPCRs and uses thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a representation of 8XCRE-Luc reporter plasmid (see, Example 4(c)(3).)

Figures 2A and 2B are graphic representations of the results of ATP and ADP binding to endogenous TDAG8 (2A) and comparisons in serum and serum free media (2B).

Figure 3 is a graphic representation of the comparative signaling results of CMV versus the GPCR Fusion Protein H9(F236K):Gsa.

DETAILED DESCRIPTION

The scientific literature that has evolved around receptors has adopted a number of terms to refer to ligands having various effects on receptors. For clarity and consistency, the following definitions will be used throughout this patent document. To the extent that these definitions conflict with other definitions for these terms, the following definitions shall control:

AGONISTS shall mean materials (e.g., ligands, candidate compounds) that

activate the intracellular response when they bind to the receptor, or enhance GTP binding to membranes.

AMINO ACID ABBREVIATIONS used herein are set out in Table A:

TABLE A			
5	ALANINE	ALA	A
	ARGININE	ARG	R
	ASPARAGINE	ASN	N
	ASPARTIC ACID	ASP	D
10	CYSTEINE	CYS	C
	GLUTAMIC ACID	GLU	E
	GLUTAMINE	GLN	Q
	GLYCINE	GLY	G
	HISTIDINE	HIS	H
15	ISOLEUCINE	ILE	I
	LEUCINE	LEU	L
	LYSINE	LYS	K
	METHIONINE	MET	M
	PHENYLALANINE	PHE	F
20	PROLINE	PRO	P
	SERINE	SER	S
	THREONINE	THR	T
	TRYPTOPHAN	TRP	W
	TYROSINE	TYR	Y
	VALINE	VAL	V

25 **PARTIAL AGONISTS** shall mean materials (e.g., ligands, candidate compounds) that activate the intracellular response when they bind to the receptor to a lesser degree/extent than do agonists, or enhance GTP binding to membranes to a lesser degree/extent than do agonists.

30 **ANTAGONIST** shall mean materials (e.g., ligands, candidate compounds) that competitively bind to the receptor at the same site as the agonists but which do not activate the intracellular response initiated by the active form of the receptor, and can thereby inhibit the intracellular responses by agonists or partial agonists. **ANTAGONISTS** do not diminish the baseline intracellular response in the absence of an agonist or partial agonist.

CANDIDATE COMPOUND shall mean a molecule (for example, and not limitation,

- 7 -

a chemical compound) that is amenable to a screening technique. Preferably, the phrase "candidate compound" does not include compounds which were publicly known to be compounds selected from the group consisting of inverse agonist, agonist or antagonist to a receptor, as previously determined by an indirect identification process ("indirectly identified compound"); more preferably, not including an indirectly identified compound which has previously been determined to have therapeutic efficacy in at least one mammal, and, most preferably, not including an indirectly identified compound which has previously been determined to have therapeutic utility in humans.

COMPOSITION means a material comprising at least one component; a "pharmaceutical composition" is an example of a composition.

COMPOUND EFFICACY shall mean a measurement of the ability of a compound to inhibit or stimulate receptor functionality, as opposed to receptor binding affinity. Exemplary means of detecting compound efficacy are disclosed in the Example section of this patent document.

CODON shall mean a grouping of three nucleotides (or equivalents to nucleotides) which generally comprise a nucleoside (adenosine (A), guanosine (G), cytidine (C), uridine (U) and thymidine (T)) coupled to a phosphate group and which, when translated, encodes an amino acid.

CONSTITUTIVELY ACTIVATED RECEPTOR shall mean a receptor subject to constitutive receptor activation. A constitutively activated receptor can be endogenous or non-endogenous.

CONSTITUTIVE RECEPTOR ACTIVATION shall mean stabilization of a receptor in the active state by means other than binding of the receptor with its endogenous

- 8 -

ligand or a chemical equivalent thereof.

CONTACT or CONTACTING shall mean bringing at least two moieties together, whether in an in vitro system or an in vivo system.

DIRECTLY IDENTIFYING or DIRECTLY IDENTIFIED, in relationship to the phrase "candidate compound", shall mean the screening of a candidate compound against a constitutively activated receptor, preferably a constitutively activated orphan receptor, and most preferably against a constitutively activated G protein-coupled cell surface orphan receptor, and assessing the compound efficacy of such compound. This phrase is, under no circumstances, to be interpreted or understood to be encompassed by or to encompass the phrase "indirectly identifying" or "indirectly identified."

ENDOGENOUS shall mean a material that a mammal naturally produces. **ENDOGENOUS** in reference to, for example and not limitation, the term "receptor," shall mean that which is naturally produced by a mammal (for example, and not limitation, a human) or a virus. By contrast, the term **NON-ENDOGENOUS** in this context shall mean that which is not naturally produced by a mammal (for example, and not limitation, a human) or a virus. For example, and not limitation, a receptor which is not constitutively active in its endogenous form, but when manipulated becomes constitutively active, is most preferably referred to herein as a "non-endogenous, constitutively activated receptor." Both terms can be utilized to describe both "in vivo" and "in vitro" systems. For example, and not limitation, in a screening approach, the endogenous or non-endogenous receptor may be in reference to an in vitro screening system. As a further example and not limitation, where the genome of a mammal has been manipulated to include a non-endogenous constitutively activated receptor, screening of a candidate compound by means of an in vivo system is viable.

G PROTEIN COUPLED RECEPTOR FUSION PROTEIN and GPCR FUSION

PROTEIN, in the context of the invention disclosed herein, each mean a non-endogenous protein comprising an endogenous, constitutively activate GPCR or a non-endogenous, constitutively activated GPCR fused to at least one G protein, most preferably the alpha (α) subunit of such G protein (this being the subunit that binds GTP), with the G protein preferably being of the same type as the G protein that naturally couples with endogenous orphan GPCR. For example, and not limitation, in an endogenous state, if the G protein "Gsa" is the predominate G protein that couples with the GPCR, a GPCR Fusion Protein based upon the specific GPCR would be a non-endogenous protein comprising the GPCR fused to Gsa; in some circumstances, as will be set forth below, a non-predominant G protein can be fused to the GPCR. The G protein can be fused directly to the c-terminus of the constitutively active GPCR or there may be spacers between the two.

HOST CELL shall mean a cell capable of having a Plasmid and/or Vector incorporated therein. In the case of a prokaryotic Host Cell, a Plasmid is typically replicated as a autonomous molecule as the Host Cell replicates (generally, the Plasmid is thereafter isolated for introduction into a eukaryotic Host Cell); in the case of a eukaryotic Host Cell, a Plasmid is integrated into the cellular DNA of the Host Cell such that when the eukaryotic Host Cell replicates, the Plasmid replicates. Preferably, for the purposes of the invention disclosed herein, the Host Cell is eukaryotic, more preferably, mammalian, and most preferably selected from the group consisting of 293, 293T and COS-7 cells.

INDIRECTLY IDENTIFYING or INDIRECTLY IDENTIFIED means the traditional approach to the drug discovery process involving identification of an endogenous ligand specific for an endogenous receptor, screening of candidate compounds against the

receptor for determination of those which interfere and/or compete with the ligand-receptor interaction, and assessing the efficacy of the compound for affecting at least one second messenger pathway associated with the activated receptor.

INHIBIT or INHIBITING, in relationship to the term "response" shall mean that a response is decreased or prevented in the presence of a compound as opposed to in the absence of the compound.

INVERSE AGONISTS shall mean materials (e.g., ligand, candidate compound) which bind to either the endogenous form of the receptor or to the constitutively activated form of the receptor, and which inhibit the baseline intracellular response initiated by the active form of the receptor below the normal base level of activity which is observed in the absence of agonists or partial agonists, or decrease GTP binding to membranes. Preferably, the baseline intracellular response is inhibited in the presence of the inverse agonist by at least 30%, more preferably by at least 50%, and most preferably by at least 75%, as compared with the baseline response in the absence of the inverse agonist.

KNOWN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has been identified.

LIGAND shall mean an endogenous, naturally occurring molecule specific for an endogenous, naturally occurring receptor.

MUTANT or MUTATION in reference to an endogenous receptor's nucleic acid and/or amino acid sequence shall mean a specified change or changes to such endogenous sequences such that a mutated form of an endogenous, non-constitutively activated receptor evidences constitutive activation of the receptor. In terms of equivalents to specific sequences, a subsequent mutated form of a human receptor is considered to be equivalent to

- 11 -

a first mutation of the human receptor if (a) the level of constitutive activation of the subsequent mutated form of a human receptor is substantially the same as that evidenced by the first mutation of the receptor; and (b) the percent sequence (amino acid and/or nucleic acid) homology between the subsequent mutated form of the receptor and the first mutation of the receptor is at least about 80%, more preferably at least about 90% and most preferably at least 95%. Ideally, and owing to the fact that the most preferred cassettes disclosed herein for achieving constitutive activation includes a single amino acid and/or codon change between the endogenous and the non-endogenous forms of the GPCR, the percent sequence homology should be at least 98%.

NON-ORPHAN RECEPTOR shall mean an endogenous naturally occurring molecule specific for an endogenous naturally occurring ligand wherein the binding of a ligand to a receptor activates an intracellular signaling pathway.

ORPHAN RECEPTOR shall mean an endogenous receptor for which the endogenous ligand specific for that receptor has not been identified or is not known.

PHARMACEUTICAL COMPOSITION shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, and not limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

PLASMID shall mean the combination of a Vector and cDNA. Generally, a Plasmid is introduced into a Host Cell for the purposes of replication and/or expression of the cDNA as a protein.

- 12 -

STIMULATE or STIMULATING, in relationship to the term "response" shall mean that a response is increased in the presence of a compound as opposed to in the absence of the compound.

VECTOR in reference to cDNA shall mean a circular DNA capable of incorporating at least one cDNA and capable of incorporation into a Host Cell.

The order of the following sections is set forth for presentational efficiency and is not intended, nor should be construed, as a limitation on the disclosure or the claims to follow.

A. Introduction

The traditional study of receptors has always proceeded from the a priori assumption (historically based) that the endogenous ligand must first be identified before discovery could proceed to find antagonists and other molecules that could affect the receptor. Even in cases where an antagonist might have been known first, the search immediately extended to looking for the endogenous ligand. This mode of thinking has persisted in receptor research even after the discovery of constitutively activated receptors. What has not been heretofore recognized is that it is the active state of the receptor that is most useful for discovering agonists, partial agonists, and inverse agonists of the receptor. For those diseases which result from an overly active receptor or an under-active receptor, what is desired in a therapeutic drug is a compound which acts to diminish the active state of a receptor or enhance the activity of the receptor, respectively, not necessarily a drug which is an antagonist to the endogenous ligand. This is because a compound that reduces or enhances the activity of the active receptor state need not bind at the same site as the endogenous ligand. Thus, as taught by a method of this invention, any search for therapeutic compounds should start by screening compounds against the ligand-independent active state.

- 13 -

B. Identification of Human GPCRs

The efforts of the Human Genome project has led to the identification of a plethora of information regarding nucleic acid sequences located within the human genome; it has been the case in this endeavor that genetic sequence information has been made available without an understanding or recognition as to whether or not any particular genomic sequence does or may contain open-reading frame information that translate human proteins. Several methods of identifying nucleic acid sequences within the human genome are within the purview of those having ordinary skill in the art. For example, and not limitation, a variety of human GPCRs, disclosed herein, were discovered by reviewing the GenBank™ database, while other GPCRs were discovered by utilizing a nucleic acid sequence of a GPCR, previously sequenced, to conduct a BLAST™ search of the EST database. Table B, below, lists several endogenous GPCRs that we have discovered, along with a GPCR's respective homologous receptor.

TABLE B

Disclosed Human Orphan GPCRs	Accession Number Identified	Open Reading Frame (Base Pairs)	Per Cent Homology To Designated GPCR	Reference To Homologous GPCR (Accession No.)
hARE-3	AL033379	1,260 bp	52.3% LPA-R	U92642
hARE-4	AC006087	1,119 bp	36% P2Y5	AF000546
hARE-5	AC006255	1,104 bp	32% <i>Oryzias latipes</i>	D43633
hGPR27	AA775870	1,128 bp	43% KIAA0001	D13626
hARE-1	AI090920	999 bp	53% GPR27	
hARE-2	AA359504	1,122 bp	39% EBI1	L31581
hPPRI	H67224	1,053 bp	31% GPR4	L36148
hG2A	AA754702	1,113 bp		

- 14 -

hRUP3	AL035423	1,005 bp	30% <i>Drosophila melanogaster</i>	2133653
hRUP4	AI307658	1,296 bp	32% pNPGPR and 29% <i>Zebrafish</i> Ya and Yb, AAB94616	NP_004876 AAC41276
hRUP5	AC005849	1,413 bp	25% DEZ	Q99788
hRUP6	AC005871	1,245 bp	23% FMLPR	P21462
hRUP7	AC007922	1,173 bp	48% GPR66	NP_006047
hCHN3	EST 36581	1,113 bp	43% H3R	AF140538
hCHN4	AA804531	1,077 bp	53% GPR27	
hCHN6	EST 2134670	1,503 bp	32% thrombin	4503637
hCHN8	EST 764455	1,029 bp	36% edg-1	NP_001391
hCHN9	EST 1541536	1,077 bp	47% KIAA0001	D13626
hCHN10	EST 1365839	1,055 bp	41% LTB4R	NM_000752
			35% P2Y	NM_002563

Receptor homology is useful in terms of gaining an appreciation of a role of the receptors within the human body. As the patent document progresses, we will disclose techniques for mutating these receptors to establish non-endogenous, constitutively activated versions of these receptors.

The techniques disclosed herein have also been applied to other human, orphan GPCRs known to the art, as will be apparent as the patent document progresses.

C. Receptor Screening

Screening candidate compounds against a non-endogenous, constitutively activated version of the human GPCRs disclosed herein allows for the direct identification of candidate compounds which act at this cell surface receptor, without requiring use of the receptor's endogenous ligand. By determining areas within the body where the endogenous version of human GPCRs disclosed herein is expressed and/or over-expressed, it is possible to determine related disease/disorder states which are associated with the expression and/or over-expression

- 15 -

of the receptor, such an approach is disclosed in this patent document.

With respect to creation of a mutation that may evidence constitutive activation of the human GPCR disclosed herein is based upon the distance from the proline residue at which is presumed to be located within TM6 of the GPCR; this algorithmic technique is disclosed in co-pending and commonly assigned patent document U.S. Serial Number 09/170,496, incorporated herein by reference. The algorithmic technique is not predicated upon traditional sequence "alignment" but rather a specified distance from the aforementioned TM6 proline residue. By mutating the amino acid residue located 16 amino acid residues from this residue, (presumably located in the IC3 region of the receptor) to, most preferably, a lysine residue, such activation may be obtained. Other amino acid residues may be useful in the mutation at this position to achieve this objective.

D. Disease/Disorder Identification and/or Selection

As will be set forth in greater detail below, most preferably inverse agonists to the non-endogenous, constitutively activated GPCR can be identified by the methodologies of this invention. Such inverse agonists are ideal candidates as lead compounds in drug discovery programs for treating diseases related to this receptor. Because of the ability to directly identify inverse agonists to the GPCR, thereby allowing for the development of pharmaceutical compositions, a search for diseases and disorders associated with the GPCR is relevant. For example, scanning both diseased and normal tissue samples for the presence of the GPCR now becomes more than an academic exercise or one which might be pursued along the path of identifying an endogenous ligand to the specific GPCR. Tissue scans can be conducted across a broad range of healthy and diseased tissues. Such tissue scans provide a preferred first step in associating a specific receptor with a disease and/or disorder. *See, for*

- 16 -

example, co-pending application (docket number ARE-0050) for exemplary dot-blot and RT-PCR results of several of the GPCRs disclosed herein.

Preferably, the DNA sequence of the human GPCR is used to make a probe for (a) dot-blot analysis against tissue-mRNA, and/or (b) RT-PCR identification of the expression of the receptor in tissue samples. The presence of a receptor in a tissue source, or a diseased tissue, or the presence of the receptor at elevated concentrations in diseased tissue compared to a normal tissue, can be preferably utilized to identify a correlation with a treatment regimen, including but not limited to, a disease associated with that disease. Receptors can equally well be localized to regions of organs by this technique. Based on the known functions of the specific tissues to which the receptor is localized, the putative functional role of the receptor can be deduced.

E. Screening of Candidate Compounds

1. Generic GPCR screening assay techniques

When a G protein receptor becomes constitutively active, it binds to a G protein (e.g., G_q, G_s, G_i, G_z, G_o) and stimulates the binding of GTP to the G protein. The G protein then acts as a GTPase and slowly hydrolyzes the GTP to GDP, whereby the receptor, under normal conditions, becomes deactivated. However, constitutively activated receptors continue to exchange GDP to GTP. A non-hydrolyzable analog of GTP, [³⁵S]GTP γ S, can be used to monitor enhanced binding to membranes which express constitutively activated receptors. It is reported that [³⁵S]GTP γ S can be used to monitor G protein coupling to membranes in the absence and presence of ligand. An example of this monitoring, among other examples well-known and available to those in the art, was reported by Traynor and Nalowski in 1995. The preferred use of this assay system is for initial screening of candidate compounds because the

system is generically applicable to all G protein-coupled receptors regardless of the particular G protein that interacts with the intracellular domain of the receptor.

2. Specific GPCR screening assay techniques

Once candidate compounds are identified using the "generic" G protein-coupled receptor assay (*i.e.*, an assay to select compounds that are agonists, partial agonists, or inverse agonists), further screening to confirm that the compounds have interacted at the receptor site is preferred. For example, a compound identified by the "generic" assay may not bind to the receptor, but may instead merely "uncouple" the G protein from the intracellular domain.

a. Gs, Gz and Gi.

Gs stimulates the enzyme adenylyl cyclase. Gi (and Gz and Go), on the other hand, inhibit this enzyme. Adenylyl cyclase catalyzes the conversion of ATP to cAMP; thus, constitutively activated GPCRs that couple the Gs protein are associated with increased cellular levels of cAMP. On the other hand, constitutively activated GPCRs that couple Gi (or Gz, Go) protein are associated with decreased cellular levels of cAMP. *See, generally,*

"Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Thus, assays that detect cAMP can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to the receptor (*i.e.*, such a compound would decrease the levels of cAMP). A variety of approaches known in the art for measuring cAMP can be utilized; a most preferred approach relies upon the use of anti-cAMP antibodies in an ELISA-based format. Another type of assay that can be utilized is a whole cell second messenger reporter system assay. Promoters on genes drive the expression of the proteins that a particular gene encodes. Cyclic AMP drives gene expression by promoting the binding of a cAMP-responsive DNA binding protein or

transcription factor (CREB) that then binds to the promoter at specific sites called cAMP response elements and drives the expression of the gene. Reporter systems can be constructed which have a promoter containing multiple cAMP response elements before the reporter gene, *e.g.*, β -galactosidase or luciferase. Thus, a constitutively activated Gs-linked receptor causes the accumulation of cAMP that then activates the gene and expression of the reporter protein. The reporter protein such as β -galactosidase or luciferase can then be detected using standard biochemical assays (Chen et al. 1995).

b. Go and Gq.

Gq and Go are associated with activation of the enzyme phospholipase C, which in turn hydrolyzes the phospholipid PIP₂, releasing two intracellular messengers: diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP₃). Increased accumulation of IP₃ is associated with activation of Gq- and Go-associated receptors. *See, generally,* "Indirect Mechanisms of Synaptic Transmission," Chpt. 8, From Neuron To Brain (3rd Ed.) Nichols, J.G. et al eds. Sinauer Associates, Inc. (1992). Assays that detect IP₃ accumulation can be utilized to determine if a candidate compound is, *e.g.*, an inverse agonist to a Gq- or Go-associated receptor (*i.e.*, such a compound would decrease the levels of IP₃). Gq-associated receptors can also be examined using an AP1 reporter assay in that Gq-dependent phospholipase C causes activation of genes containing AP1 elements; thus, activated Gq-associated receptors will evidence an increase in the expression of such genes, whereby inverse agonists thereto will evidence a decrease in such expression, and agonists will evidence an increase in such expression. Commercially available assays for such detection are available.

3. GPCR Fusion Protein

The use of an endogenous, constitutively activate orphan GPCR or a non-endogenous, constitutively activated orphan GPCR, for use in screening of candidate compounds for the direct identification of inverse agonists, agonists and partial agonists provide an interesting screening challenge in that, by definition, the receptor is active even in the absence of an endogenous ligand bound thereto. Thus, in order to differentiate between, e.g., the non-endogenous receptor in the presence of a candidate compound and the non-endogenous receptor in the absence of that compound, with an aim of such a differentiation to allow for an understanding as to whether such compound may be an inverse agonist, agonist, partial agonist or have no effect on such a receptor, it is preferred that an approach be utilized that can enhance such differentiation. A preferred approach is the use of a GPCR Fusion Protein.

Generally, once it is determined that a non-endogenous orphan GPCR has been constitutively activated using the assay techniques set forth above (as well as others), it is possible to determine the predominant G protein that couples with the endogenous GPCR.

15 Coupling of the G protein to the GPCR provides a signaling pathway that can be assessed. Because it is most preferred that screening take place by use of a mammalian expression system, such a system will be expected to have endogenous G protein therein. Thus, by definition, in such a system, the non-endogenous, constitutively activated orphan GPCR will continuously signal. In this regard, it is preferred that this signal be enhanced such that in the presence of, e.g., an inverse agonist to the receptor, it is more likely that it will be able to more readily differentiate, particularly in the context of screening, between the receptor when it is contacted with the inverse agonist.

The GPCR Fusion Protein is intended to enhance the efficacy of G protein coupling

with the non-endogenous GPCR. The GPCR Fusion Protein is preferred for screening with a non-endogenous, constitutively activated GPCR because such an approach increases the signal that is most preferably utilized in such screening techniques. This is important in facilitating a significant "signal to noise" ratio; such a significant ratio is import preferred for the screening of candidate compounds as disclosed herein.

5 The construction of a construct useful for expression of a GPCR Fusion Protein is within the purview of those having ordinary skill in the art. Commercially available expression vectors and systems offer a variety of approaches that can fit the particular needs of an investigator. The criteria of importance for such a GPCR Fusion Protein construct is that the endogenous GPCR sequence and the G protein sequence both be in-frame (preferably), the sequence for the endogenous GPCR is upstream of the G protein sequence) and that the "stop" codon of the GPCR must be deleted or replaced such that upon expression of the GPCR, the G protein can also be expressed. The GPCR can be linked directly to the G protein, or there can be spacer residues between the two (preferably, no more than about 12. 10 although this number can be readily ascertained by one of ordinary skill in the art). We have a preference (based upon convenience) of use of a spacer in that some restriction sites that are not used will, effectively, upon expression, become a spacer. Most preferably, the G protein that couples to the non-endogenous GPCR will have been identified prior to the creation of the GPCR Fusion Protein construct. Because there are only a few G proteins that have been 20 identified, it is preferred that a construct comprising the sequence of the G protein (i.e., a universal G protein construct) be available for insertion of an endogenous GPCR sequence therein; this provides for efficiency in the context of large-scale screening of a variety of different endogenous GPCRs having different sequences.

As noted above, constitutively activated GPCRs that couple to Gi, Gz and Go are expected to inhibit the formation of cAMP making assays based upon these types of GPCRs challenging (*i.e.*, the cAMP signal decreases upon activation thus making the direct identification of, *e.g.* inverse agonists (which would further decrease this signal), interesting).

5 As will be disclosed herein, we have ascertained that for these types of receptors, it is possible to create a GPCR Fusion Protein that is not based upon the endogenous GPCR's endogenous G protein, in an effort to establish a viable cyclase-based assay. Thus, for example, a Gz coupled receptor such as H9, a GPCR Fusion Protein can be established that utilizes a Gs fusion protein - we believe that such a fusion construct, upon expression, "drives" or "forces" the non-endogenous GPCR to couple with, *e.g.*, Gs rather than the "natural" Gz protein, such that a cyclase-based assay can be established. Thus, for Gi, Gz and Go coupled receptors, we prefer that that when a GPCR Fusion Protein is used and the assay is based upon detection of adenylyl cyclase activity, that the fusion construct be established with Gs (or an equivalent G protein that stimulates the formation of the enzyme adenylyl cyclase).

15 F. Medicinal Chemistry

Generally, but not always, direct identification of candidate compounds is preferably conducted in conjunction with compounds generated via combinatorial chemistry techniques, whereby thousands of compounds are randomly prepared for such analysis. Generally, the results of such screening will be compounds having unique core structures; thereafter, these compounds are preferably subjected to additional chemical modification around a preferred core structure(s) to further enhance the medicinal properties thereof. Such techniques are known to those in the art and will not be addressed in detail in this patent document.

G. Pharmaceutical compositions

Candidate compounds selected for further development can be formulated into pharmaceutical compositions using techniques well known to those in the art. Suitable pharmaceutically-acceptable carriers are available to those in the art; for example, see Remington's Pharmaceutical Sciences, 16th Edition, 1980, Mack Publishing Co., (Oslo et al., eds.)

H. Other Utility

Although a preferred use of the non-endogenous versions the human GPCRs disclosed herein may be for the direct identification of candidate compounds as inverse agonists, agonists or partial agonists (preferably for use as pharmaceutical agents), these versions of human GPCRs can also be utilized in research settings. For example, *in vitro* and *in vivo* systems incorporating GPCRs can be utilized to further elucidate and understand the roles these receptors play in the human condition, both normal and diseased, as well as understanding the role of constitutive activation as it applies to understanding the signaling cascade. The value in non-endogenous human GPCRs is that their utility as a research tool is enhanced in that, because of their unique features, non-endogenous human GPCRs can be used to understand the role of these receptors in the human body before the endogenous ligand therefor is identified. Other uses of the disclosed receptors will become apparent to those in the art based upon, *inter alia*, a review of this patent document.

20

EXAMPLES

The following examples are presented for purposes of elucidation, and not limitation, of the present invention. While specific nucleic acid and amino acid sequences are disclosed herein, those of ordinary skill in the art are credited with the ability to make minor

- 23 -

modifications to these sequences while achieving the same or substantially similar results reported below. The traditional approach to application or understanding of sequence cassettes from one sequence to another (e.g. from rat receptor to human receptor or from human receptor A to human receptor B) is generally predicated upon sequence alignment techniques whereby the sequences are aligned in an effort to determine areas of commonality. The mutational approach disclosed herein does not rely upon this approach but is instead based upon an algorithmic approach and a positional distance from a conserved proline residue located within the TM6 region of human GPCRs. Once this approach is secured, those in the art are credited with the ability to make minor modifications hereto to achieve substantially the same results (*i.e.*, constitutive activation) disclosed herein. Such modified approaches are considered within the purview of this disclosure.

Example 1 ENDOGENOUS HUMAN GPCRS

1. Identification of Human GPCRS

Certain of the disclosed endogenous human GPCRS were identified based upon a review of the GenBank™ database information. While searching the database, the following cDNA clones were identified as evidenced below (Table C).

TABLE C

Disclosed Human Orphan GPCRS	Accession Number	Complete DNA Sequence (Base Pairs)	Open Reading Frame (Base Pairs)	Nucleic Acid SEQ.ID. NO.	Amino Acid SEQ.ID. NO.
hARE-3	AL033379	111,389 bp	1,260 bp	1	2
hARE-4	AC006087	226,925 bp	1,119 bp	3	4
hARE-5	AC006235	127,605 bp	1,104 bp	5	6
hRUP3	AL035423	140,094 bp	1,005 bp	7	8

- 24 -

hRUP5	AC005849	169,144 bp	1,413 bp	9	10
hRUP6	AC005871	218,807 bp	1,245 bp	11	12
hRUP7	AC007922	158,858 bp	1,173 bp	13	14

Other disclosed endogenous human GPCRS were identified by conducting a BLAST™ search of EST database (dbest) using the following EST clones as query sequences. The following EST clones identified were then used as a probe to screen a human genomic library (Table D).

TABLE D

Disclosed Human Orphan GPCRS	Query (Sequence)	EST Clone/Accession No. Identified	Open Reading Frame (Base Pairs)	Nucleic Acid SEQ.ID. NO.	Amino Acid SEQ.ID. NO.
hGPCR27	Mouse	AA775870	1,125 bp	17	18
hARE-1	GPCR27	1689643	999 bp	19	20
hGPCR27	TDAG	A1090920	1,122 bp	21	22
hARE-2	GPCR27	68530	1,053 bp	23	24
hPPR1	Bovine	238667	1,113 bp	25	26
hG2A	Mouse	H67224	1,113 bp	27	28
hCHN3	Mouse	See Example 2(a), below	1,113 bp	29	30
hCHN4	TDAG	1184934	1,077 bp	31	32
hCHN6	N.A.	AA804531	1,503 bp	33	34
hCHN8	K1AA0001	EST 2134670	1,029 bp	35	36
hCHN9	1365839	EST 764455	1,077 bp	37	38
hCHN10	Mouse EST	EST 1541536	1,005 bp	39	40
hRUP4	Human	1365839	1,296 bp	39	40

N.A. = "not applicable".

2. Full Length Cloning

a. Human G2A

Mouse EST clone 1179426 was used to obtain a human genomic clone containing all

but three amino acid G2A coding sequences. The 5' of this coding sequence was obtained by using 5'RACE, and the template for PCR was Clontech's Human Spleen Marathon-Ready™ cDNA. The disclosed human G2A was amplified by PCR using the G2A cDNA specific primers for the first and second round PCR as shown in SEQ.ID.NO.: 41 and SEQ.ID.NO.: 42 as follows:

5' - CTGTGTACAGCAGTTCCGACAGTG-3' (SEQ.ID.NO.: 41; 1st round PCR)

5' - GAGTGCCAGGCAGAGCGTAGAC-3' (SEQ.ID.NO.: 42; second round PCR).

PCR was performed using Advantage GC Polymerase Kit (Clontech; manufacturing instructions will be followed), at 94°C for 30 sec followed by 5 cycles of 94°C for 5 sec and 72°C for 4 min; and 30 cycles of 94° for 5 sec and 70° for 4 min. An approximate 1.3 Kb PCR fragment was purified from agarose gel, digested with Hind III and Xba I and cloned into the expression vector pRC/CMV2 (Invitrogen). The cloned-insert was sequenced using the T7 Sequenase™ kit (USB Amersham; manufacturer instructions followed) and the sequence was compared with the presented sequence. Expression of the human G2A was detected by probing an RNA dot blot (Clontech; manufacturer instructions followed) with the P³²-labeled fragment.

b. CHN9

Sequencing of the EST clone 1541536 showed CHN9 to be a partial cDNA clone having only an initiation codon; *i.e.*, the termination codon was missing. When CHN9 was used to blast against data base (nr), the 3' sequence of CHN9 was 100% homologous to the 5' untranslated region of the leukotriene B4 receptor cDNA, which contained a termination codon in the frame with CHN9 coding sequence. To determine whether the 5' untranslated region of LTB4R cDNA was the 3' sequence of CHN9, PCR was performed using primers based upon the 5' sequence flanking the initiation codon found in CHN9 and

the 3' sequence around the termination codon found in the LTB4R 5' untranslated region. The 5' primer sequence utilized was as follows:

5' - CCCGAATTCCTGCTGTCCAGCTTGCCCC-3' (SEQ.ID.NO.: 43; sense) and

5' - TGTGGATCCTGCTGTCAAAAGGTCCCATTCGGG-3' (SEQ.ID.NO.: 44; antisense).

5 PCR was performed using thymus cDNA as a template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 uM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 65°C for 1 min and 72°C for 1 min and 10 sec. A 1.1kb fragment consistent with the predicted size was obtained from PCR. This PCR fragment was subcloned into pCMV (*see below*) and sequenced (*see, SEQ.ID.NO.: 35*).

c. RUP 4

The full length RUP4 was cloned by RT-PCR with human brain cDNA (Clontech) as templates:

5' - TCACAAATGCTAGGTGTGGTC-3' (SEQ.ID.NO.: 45; sense) and

15 5' - TGCATAGACAAATGGGATTACAG-3' (SEQ.ID.NO.: 46; antisense).

PCR was performed using TaqPlus Precision™ polymerase (Stratagene; manufacturing instructions followed) by the following cycles: 94°C for 2 min; 94°C 30 sec; 55°C for 30 sec, 72°C for 45 sec, and 72°C for 10 min. Cycles 2 through 4 were repeated 30 times.

The PCR products were separated on a 1% agarose gel and a 500 bp PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and sequenced using the T7 DNA Sequenase™ kit (Amsham) and the SP6/T7 primers (Stratagene). Sequence analysis revealed that the PCR fragment was indeed an alternatively spliced form of A1307658 having a continuous open reading frame with similarity to other GPCRs. The completed sequence of this PCR fragment was as follows:

- 27 -

5'-TCACAATGCTAGTGTGGTGGCTGGGACATCATCTAGAGATCACCATTGTGGCAC
 GTGCAACAATTGAGATCAAAATATGACTCTCATATGAAAAGAAACACATCTGCTTAAGA
 GTGGACCGACCTGTGCACCGAAGATCTACACCTTCATCTCTGCTCTCTCTG
 CTCTTATGTGTGATGCTTATCTGTAGTAATAATTGGTATGAACTTGGATATAAGAAAGAGT
 GGGGATGTTCAAGTCTTCCAACTATTGACAGTGGTGGCTCTTCTGCTGTGGGACCATTC
 AAACGAGCTGCTATTATGATGTGATGATACGATAATTGAAAAGAAATATGATGATGATCAATCAA
 GATGATTTTGTCTATCGCAAAATTAATTGATTTCCAACTCATCTGTAAATCCCATTTGCTATGCA-
 3' (SEQ.ID.NO.: 47)

10 Based on the above sequence, two sense oligonucleotide primer sets:

5'-CTGCTTAGAAGAGTGACACCAG-3' (SEQ.ID.NO.: 48; oligo 1),

5'-CTGTGCACCGAAGATCTACAC-3' (SEQ.ID.NO.: 49; oligo 2) and

two antisense oligonucleotide primer sets:

5'-CAAGGATGAAGGTGGTGTAGA-3' (SEQ.ID.NO.: 50; oligo 3)

15 5'-GTGATGATCTTCTGTGCACAGG-3' (SEQ.ID.NO.: 51; oligo 4)

were used for 3' - and 5'-RACE PCR with a human brain Marathon-Ready™ cDNA
 (Clontech, Cat# 7400-1) as template, according to manufacture's instructions. DNA
 fragments generated by the RACE PCR were cloned into the pCRII-TOPO™ vector
 (Invitrogen) and sequenced using the SP6/T7 primers (Stratagene) and some internal primers.

20 The 3' RACE product contained a poly(A) tail and a completed open reading frame ending
 at a TAA stop codon. The 5' RACE product contained an incomplete 5' end, i.e., the ATG
 initiation codon was not present.

Based on the new 5' sequence, oligo 3 and the following primer:

5'-GCAATGACAGTGCATAGTGAGC-3' (SEQ.ID.NO.: 52; oligo 5)

25 were used for the second round of 5' race PCR and the PCR products were analyzed as above.
 A third round of 5' race PCR was carried out utilizing antisense primers:

5'-TGAGACATGCTGTGACGGGAATGCAGAG-3' (SEQ.ID.NO.: 53; oligo 6) and

5'-GTGATGACAGTGCATGAGGCCAAG-3' (SEQ.ID.NO.: 54; oligo 7).

The sequence of the 5' RACE PCR products revealed the presence of the initiation codon

- 28 -

ATG, and further round of 5' race PCR did not generate any more 5' sequence. The
 completed 5' sequence was confirmed by RT-PCR using sense primer

5'-GCAATGACAGCGCTTAACATTAC-3' (SEQ.ID.NO.: 55; oligo 8)

and oligo 4 as primers and sequence analysis of the 650 bp PCR product generated from
 5 human brain and heart cDNA templates (Clontech, Cat# 7404-1). The completed 3' sequence
 was confirmed by RT-PCR using oligo 2 and the following antisense primer:

5'-TTGGGTTACCAATCTGAAGGCA-3' (SEQ.ID.NO.: 56; oligo 9)

and sequence analysis of the 670 bp PCR product generated from human brain and heart
 cDNA templates (Clontech, Cat# 7404-1).

10 d. RUP5

The full length RUP5 was cloned by RT-PCR using a sense primer upstream from
 ATG, the initiation codon (SEQ.ID.NO.: 57), and an antisense primer containing TCA as the
 stop codon (SEQ.ID.NO.: 58), which had the following sequences:

5'-ACTCCGTGTCACGACGAGACTGTG-3' (SEQ.ID.NO.: 57)

15 5'-TGGGTGTTCTGGAACCTCAGTG-3' (SEQ.ID.NO.: 58)

and human peripheral leukocyte cDNA (Clontech) as a template. Advantage™ cDNA
 polymerase (Clontech) was used for the amplification in a 50ul reaction by the following cycle
 with step 2 through step 4 repeated 30 times: 94° C for 30 sec; 94° for 15 sec; 69° for 40 sec;
 72° C for 3 min; and 72° C for 6 min. A 1.4kb PCR fragment was isolated and cloned with
 20 the pCRII-TOPO™ vector (Invitrogen) and completely sequenced using the T7 DNA
 Sequenase™ kit (Amsham). See, SEQ.ID.NO.: 9.

e. RUP6

The full length RUP6 was cloned by RT-PCR using primers:

5'-CAGGCCCTGGATTTAATGTCAGGAGATGG-3' (SEQ.ID.NO.: 59) and

5'-GGAGAGTCAGCTCTGAAAGAAATTCAGG-3' (SEQ.ID.NO.: 60);

and human thymus Marathon-Ready™ cDNA (Clontech) as a template. Advantage cDNA polymerase (Clontech, according to manufacturer's instructions) was used for the amplification in a 50ul reaction by the following cycle: 94°C for 30sec; 94°C for 5 sec; 66°C for 40sec; 72°C for 2.5 sec and 72°C for 7 min. Cycles 2 through 4 were repeated 30 times.

A 1.3 Kb PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced (*see*, SEQ.ID.NO.: 11) using the ABI Big Dye Terminator™ kit (P.E. Biosystem).

f. RUP7

10 The full length RUP7 was cloned by RT-PCR using primers:

5'-TGATGTGATGCCAGATACTAATAGCAC-3' (SEQ.ID.NO.: 61; sense) and

5'-CCTGATTCAATTAGGTGAGATTGAGAC-3' (SEQ.ID.NO.: 62; antisense)

and human peripheral leukocyte cDNA (Clontech) as a template. Advantage™ cDNA polymerase (Clontech) was used for the amplification in a 50 ul reaction by the following cycle with step 2 to step 4 repeated 30 times: 94°C for 2 minutes; 94°C for 15 seconds; 60°C for 20 seconds; 72°C for 2 minutes; 72°C for 10 minutes. A 1.25 Kb PCR fragment was isolated and cloned into the pCRII-TOPO™ vector (Invitrogen) and completely sequenced using the ABI Big Dye Terminator™ kit (P.E. Biosystem). *See*, SEQ.ID.NO.: 13.

3. Angiotensin II Type 1 Receptor ("AT1")

20 The endogenous human angiotensin II type 1 receptor ("AT1") was obtained by PCR using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94°C for 1 min, 55°C for 1min and 72°C for 1.5 min. The 5' PCR primer contains a HindIII site with the sequence:

5'-CCCAAGCTTCCCGAGGTGATTTGAT-3' (SEQ.ID.NO.: 63)

and the 3' primer contains a BamHI site with the following sequence:

5'-GTTGGATCCACATAATGCAATTTCTC-3' (SEQ.ID.NO.: 64).

The resulting 1.3 kb PCR fragment was digested with HindIII and BamHI and cloned into 5 HindIII-BamHI site of pCMV expression vector. The cDNA clone was fully sequenced. Nucleic acid (SEQ.ID.NO.: 65) and amino acid (SEQ.ID.NO.: 66) sequences for human AT1 were thereafter determined and verified.

4. GPR38

To obtain GPR38, PCR was performed by combining two PCR fragments, using human genomic cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25μM of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition for each PCR reaction was 30 cycles of 94°C for 1 min, 62°C for 1min and 72°C for 2 min.

The first fragment was amplified with the 5' PCR primer that contained an end site 15 with the following sequence:

5'-ACCATGGGCGAGCCCTGGAAACGGCAGC-3' (SEQ.ID.NO.:67)

and a 3' primer having the following sequence:

5'-AGAACCCACCACGACGAGGACGGGACGGTCTGCGGTGG-3' (SEQ.ID.NO.:68).

The second PCR fragment was amplified with a 5' primer having the following sequence:

20 5'-GTCGCGCTCCTGCTGGTGGTGTCTTGCAATTATAATT-3' (SEQ.ID.NO.: 69)

and a 3' primer that contained a BamHI site and having the following sequence:

5'-CCTGGATCCTTATCCCATCGTCTTCACGTTAGC-3' (SEQ.ID.NO.: 70).

The two fragments were used as templates to amplify GPR38, using SEQ.ID.NO.: 67 and SEQ.ID.NO.: 70 as primers (using the above-noted cycle conditions). The resulting 1.44kb

- 31 -

PCR fragment was digested with BamHI and cloned into Blunt-BamHI site of pCMV expression vector.

5. MC4

To obtain MC4, PCR was performed using human genomic cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition for each PCR reaction was 30 cycles of 94 °C for 1 min, 54 °C for 1 min and 72 °C for 1.5 min.

The 5' PCR contained an EcoRI site with the sequence:

5'-CTGGAATTCTCTGCCAGCATGGTGA-3' (SEQ.ID.NO.: 71)

and the 3' primer contained a BamHI site with the sequence:

5'-GCAGAGATCCTATATGGCTGCTGTCCCC-3' (SEQ.ID.NO.: 72)

The 1.0 kb PCR fragment was digested with EcoRI and BamHI and cloned into EcoRI-BamHI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 73) and amino acid (SEQ.ID.NO.: 74) sequences for human MC4 were thereafter determined.

6. CCKB

To obtain CCKB, PCR was performed using human stomach cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition for each PCR reaction was 30 cycles of 94 °C for 1 min, 65 °C for 1 min and 72 °C for 1 min and 30 sec.

The 5' PCR contained a HindIII site with the sequence:

5'-CCGAAGCTTCGAGCTGAGTAAGCGCGGCT-3' (SEQ.ID.NO.: 75)

and the 3' primer contained an EcoRI site with the sequence:

5'-GTGGAATTCAATTGGCCCTGCTCAACCCCA-3' (SEQ.ID.NO.: 76)

The resulting 1.44 kb PCR fragment was digested with HindIII and EcoRI and cloned into

- 32 -

HindIII-EcoRI site of pCMV expression vector. Nucleic acid (SEQ.ID.NO.: 77) and amino acid (SEQ.ID.NO.: 78) sequences for human CCKB were thereafter determined.

7. TDAG8

To obtain TDAG8, PCR was performed using genomic DNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94 °C for 1 min, 56 °C for 1 min and 72 °C for 1 min and 20 sec. The 5' PCR primer contained a HindIII site with the following sequence:

5'-TGCAAGCTTAAAAAGGAAAAATGACAGC-3' (SEQ.ID.NO.: 79)

and the 3' primer contained a BamHI site with the following sequence:

5'-TAAAGATCCCTTCCTTCAAAACATCTTG-3' (SEQ.ID.NO.: 80)

The resulting 1.1 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. Three resulting clones sequenced contained three potential polymorphisms involving changes of amino acid 43 from Pro to Ala, amino acid 97 from Lys to Asn and amino acid 130 from Ile to Phe. Nucleic acid (SEQ.ID.NO.: 81) and amino acid (SEQ.ID.NO.: 82) sequences for human TDAG8 were thereafter determined.

8. H9

To obtain H9, PCR was performed using pituitary cDNA as template and rTth polymerase (Perkin Elmer) with the buffer system provided by the manufacturer, 0.25 μ M of each primer, and 0.2 mM of each 4 nucleotides. The cycle condition was 30 cycles of 94 °C for 1 min, 62 °C for 1 min and 72 °C for 2 min. The 5' PCR primer contained a HindIII site with the following sequence:

5'-GGAAGCTTAACGATCCCGAGAGCAACAT-3' (SEQ.ID.NO.: 15)

and the 3' primer contained a BamHI site with the following sequence:

5'-CTGGGATCTACGAGCATTTTCACACAG-3' (SEQ.ID.NO.:16).

The resulting 1.9 kb PCR fragment was digested with HindIII and BamHI and cloned into HindIII-BamHI site of pCMV expression vector. H9 contained three potential polymorphisms involving changes of amino acid P320S, S493N and amino acid G448A. Nucleic acid
 5 (SEQ.ID.NO.: 139) and amino acid (SEQ.ID.NO.: 140) sequences for human H9 were thereafter determined and verified.

Example 2

PREPARATION OF NON-ENDOGENOUS, CONSTITUTIVELY ACTIVATED GPCRS

Those skilled in the art are credited with the ability to select techniques for

10 mutation of a nucleic acid sequence. Presented below are approaches utilized to create non-endogenous versions of several of the human GPCRs disclosed above. The mutations disclosed below are based upon an algorithmic approach whereby the 16th amino acid (located in the IC3 region of the GPCR) from a conserved proline residue (located in the TM6 region of the GPCR, near the TM6/IC3 interface) is mutated, most preferably to a lysine amino acid residue.

1. Transformer Site-Directed™ Mutagenesis

Preparation of non-endogenous human GPCRs may be accomplished on human GPCRs using Transformer Site-Directed™ Mutagenesis Kit (Clontech) according to the manufacturer instructions. Two mutagenesis primers are utilized, most preferably a lysine mutagenesis oligonucleotide that creates the lysine mutation, and a selection marker oligonucleotide. For convenience, the codon mutation to be incorporated into the human GPCR is also noted, in standard form (Table E):

TABLE E

Receptor Identifier	Codon Mutation
hARE-3	F13K
hARE-4	V23K
hARE-5	A240K
hGPCR14	L257K
hGPCR27	C283K
hARE-1	E232K
hARE-2	G285K
hPPR1	L239K
hG2A	K232A
hRUP3	L224K
hRUP5	A236K
hRUP6	N267K
hRUP7	A302K
hCHN4	V236K
hMC4	A244K
hCHN3	S284K
hCHN6	L352K
hCHN8	N235K
hCHN9	G223K
hCHN10	L231K
hH9	F236K

The following GPCRs were mutated according with the above method using the designated sequence primers (Table F).

- 35 -

TABLE F

Receptor Identifier	Codon Mutation	Lysine Mutagenesis (SEQ.ID.NO.) 5'-3' orientation, mutation sequence underlined	Selection Marker (SEQ.ID.NO.) 5'-3' orientation
hRUP4	V272K	CAGGAAGAAGAAACGAGC TGTCATTATGATGTGACAC GTG (83) alternative approach, see below GGCCACCGGAGAGCCAAAC GGCTCCTGCTG (85) alternative approach, see below GGAAAGAAGAGAAATCAA AAAACACTACTGTTCAGCATC (87)	CACTGTCACCATCATTAATG ACAGCTGCTTCTTCTTCC TG (84) alternative approach, see below CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (86) alternative approach, see below CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (88)
hAT1	see below		
hGPR38	V297K	GGCCACCGGAGAGCCAAAC GGCTCCTGCTG (85) alternative approach, see below GGAAAGAAGAGAAATCAA AAAACACTACTGTTCAGCATC (87)	CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (84) CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (138)
hCCKB	V332K	GGAAAGAAGAGAAATCAA AAAACACTACTGTTCAGCATC (87)	CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (144)
hTDAG8	I225K	GGAAAGAAGAGAAATCAA AAAACACTACTGTTCAGCATC (87)	CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (144)
hH9	F236K	GCTGAGGTTTCGCAATTAAC TAACCATGTTGTG (143) GCCAATATGAAGGAGAAA ATTACCTTGACCATC (137)	GTTGTCAGAAAT (144) CTCCTTGGTCTCTCTATC GTTGTCAGAAAT (138)
hMCA	A244K		

The non-endogenous human GPCRs were then sequenced and the derived and verified nucleic acid and amino acid sequences are listed in the accompanying "Sequence Listing" appendix to this patent document, as summarized in Table G below:

TABLE G

Non Endogenous Human GPCR	Nucleic Acid Sequence Listing	Amino Acid Sequence Listing
hRUP4 (V272K)	SEQ.ID.NO.: 127	SEQ.ID.NO.: 128
hAT1 (see alternative approaches below)	(see alternative approaches below)	(see alternative approaches below)
hGPR38 (V297K)	SEQ.ID.NO.: 129	SEQ.ID.NO.: 130
hCCKB (V332K)	SEQ.ID.NO.: 131	SEQ.ID.NO.: 132
hTDAG8 (I225K)	SEQ.ID.NO.: 133	SEQ.ID.NO.: 134
hH9 (F236K)	SEQ.ID.NO.: 141	SEQ.ID.NO.: 142
hMCA (A244K)	SEQ.ID.NO.: 135	SEQ.ID.NO.: 136

- 36 -

2. Alternative Approaches For Creation of Non-Endogenous Human GPCRs

a. AT1

1. F239K Mutation

Preparation of a non-endogenous, constitutively activated human AT1 receptor was accomplished by creating an F239K mutation (see, SEQ.ID.NO.: 89 for nucleic acid sequence, and SEQ.ID.NO.: 90 for amino acid sequence). Mutagenesis was performed using Transformer Site-Directed Mutagenesis™ Kit (Clontech) according to the 10 manufacturer's instructions. The two mutagenesis primers were used, a lysine mutagenesis oligonucleotide (SEQ.ID.NO.: 91) and a selection marker oligonucleotide (SEQ.ID.NO.: 92), which had the following sequences:

5'-CCAGAAGATGATGATATTAAGAGATAATTATGGC-3' (SEQ.ID.NO.: 91)

5'-CTCCTTGGTCTCTCTATCGTTGTCAGAAAT-3' (SEQ.ID.NO.: 92),

15 respectively.

2. N111A Mutation

Preparation of a non-endogenous human AT1 receptor was also accomplished by creating an N111A mutation (see, SEQ.ID.NO.: 93 for nucleic acid sequence, and SEQ.ID.NO.: 94 for amino acid sequence). Two PCR reactions were performed using pfu polymerase (Stratagene) with the buffer system provided by the manufacturer, supplemented with 10% DMSO, 0.25 μM of each primer, and 0.5 mM of each 4 nucleotides. The 5' PCR sense primer used had the following sequence:

5'-CCCAAGCTTCCCGAGGTGATTGAT-3' (SEQ.ID.NO.: 95)

and the antisense primer had the following sequence:

5'-CTGTGAGGCGAACTGACTCTGGCTGAAG-3' (SEQ.ID.NO.: 96).

The resulting 400 bp PCR fragment was digested with HindIII site and subcloned into HindIII-SmaI site of pCMV vector (5' construct). The 3' PCR sense primer used had the following sequence:

5 5'-CTGTACGCTAGTGTGTTTCTACTACGTGTCTCAGCAATTGAT-3' (SEQ.ID.NO.: 97)

and the antisense primer had the following sequence:

5'-GTTGGATCCACATAATGCAATTTTCTC-3' (SEQ.ID.NO.: 98)

The resulting 880 bp PCR fragment was digested with BamHI and inserted into Pst (blunted by T4 polymerase) and BamHI site of 5' construct to generate the full length N111A construct. The cycle condition was 25 cycles of 94°C for 1 min, 60°C for 1 min and 72°C for 1 min (5' PCR) or 1.5 min (3' PCR).

3. AT2K255IC3 Mutation

Preparation of a non-endogenous, constitutively activated human AT1 was accomplished by creating an AT2K255IC3 "domain swap" mutation (see, SEQ.ID.NO.: 99 for nucleic acid sequence, and SEQ.ID.NO.: 100 for amino acid sequence). Restriction sites flanking IC3 of AT1 were generated to facilitate replacement of the IC3 with corresponding IC3 from angiotensin II type 2 receptor (AT2). This was accomplished by performing two PCR reactions. A 5' PCR fragment (Fragment A) encoded from the 5' untranslated region to the beginning of IC3 was generated by utilizing SEQ.ID.NO.: 63 as sense primer and the following sequence:

5'-TCCGAATCCAAATAACTTGTAGAATGATCAGAAA-3' (SEQ.ID.NO.: 101)

as antisense primer. A 3' PCR fragment (Fragment B) encoding from the end of IC3 to the 3' untranslated region was generated by using the following sequence:

5'-AGATCTTAAGAAGATAATTATGGCAATTGTGCT-3' (SEQ.ID.NO.: 102)

as sense primer and SEQ.ID.NO.: 64 as antisense primer. The PCR condition was 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min using endogenous AT1 cDNA clone as template and pfu polymerase (Stratagene), with the buffer systems provided by the manufacturer, supplemented with 10% DMSO, 0.25 µM of each primer, 5 and 0.5 mM of each 4 nucleotides. Fragment A (720 bp) was digested with HindIII and EcoRI and subcloned. Fragment B was digested with BamHI and subcloned into pCMV vector with an EcoRI site 5' to the cloned PCR fragment.

The DNA fragment (Fragment C) encoding IC3 of AT2 with a L255K mutation and containing an EcoRI cohesive end at 5' and a AflIII cohesive end at 3' was generated by annealing 2 synthetic oligonucleotides having the following sequences:

5'AATTCGAAAACACCTTACTGAAGACGAATAGCTATGGGAAGAACACAGGATAACCCGTGACCAA G-3' (sense; SEQ.ID.NO.: 103)

15 5'TTAACTCGTGCACGGTTATCCTGTCTCTCCCATAGCTATTTCGTCTTCAGT
AAGTGTTCG-3' (antisense; SEQ.ID.NO.: 104).

Fragment C was inserted in front of Fragment B through EcoRI and AflIII site. The resulting clone was then ligated with the Fragment A through the EcoRI site to generate AT1 with AT2K255IC3.

4. A243+ Mutation

Preparation of a non-endogenous human AT1 receptor was also accomplished by creating an A243+ mutation (see, SEQ.ID.NO.: 105 for nucleic acid sequence, and SEQ.ID.NO.: 106 for amino acid sequence). An A243+ mutation was constructed using the following PCR based strategy: Two PCR reactions was performed using pfu polymerase (Stratagene) with the buffer system provided by the manufacturer supplemented with 10% DMSO, 0.25 µM of each primer, and 0.5 mM of each 4 nucleotides. The 5' PCR sense primer

- 39 -

utilized had the following sequence:

5'-CCCAAGCTCCCCAGGTGTATTGAT-3' (SEQ.ID.NO.: 107)

and the antisense primer had the following sequence:

5'-AAGCACAATTGCTGCATTAATTCTTAAATAATCATC-3' (SEQ.ID.NO.: 108).

The 3' PCR sense primer utilized had the following sequence:

5'-AAGATTAATTATGGCAGCAATTGTGCTTTCTTTCTTT-3' (SEQ.ID.NO.: 109)

containing the Ala insertion and antisense primer:

5'-GTTGGATCCACATAATGCAATTTTCTC-3' (SEQ.ID.NO.: 110).

The cycle condition was 25 cycles of 94°C for 1 min, 54°C for 1min and 72°C for 1.5 min.

An aliquot of the 5' and 3' PCR were then used as co-template to perform secondary PCR using the 5' PCR sense primer and 3' PCR antisense primer. The PCR condition was the

same as primary PCR except the extension time was 2.5 min. The resulting PCR fragment was digested with HindIII and BamHI and subcloned into pCMV vector. (See,

SEQ.ID.NO.: 105)

4. CCKB

Preparation of the non-endogenous, constitutively activated human CCKB receptor

was accomplished by creating a V322K mutation (see, SEQ.ID.NO.: 111 for nucleic acid sequence and SEQ.ID.NO.: 112 for amino acid sequence). Mutagenesis was performed by

PCR via amplification using the wildtype CCKB from Example 1.

The first PCR fragment (1kb) was amplified by using SEQ.ID.NO.: 75 and an

antisense primer comprising a V322K mutation:

5'-CAGCAGCATGCGCTTACGCGCTTCTTAGCCAG-3' (SEQ.ID.NO.: 113).

The second PCR fragment (0.44kb) was amplified by using a sense primer comprising the

V322K mutation:

- 40 -

5'-AGAACGCCGCTGAAGCCCATGCTGCTGTGATCGTT-3' (SEQ.ID.NO.: 114) and SEQ.ID.NO.: 76.

The two resulting PCR fragments were then used as template for amplifying CCKB comprising V322K, using SEQ.ID.NO.: 75 and SEQ.ID.NO.: 76 and the above-noted

system and conditions. The resulting 1.44kb PCR fragment containing the V322K

mutation was digested with HindIII and EcoRI and cloned into HindIII-EcoRI site of pCMV expression vector. (See, SEQ.ID.NO.: 111).

3. QuikChange™ Site-Directed™ Mutagenesis

Preparation of non-endogenous human GPCRs can also be accomplished by using

QuikChange™ Site-Directed™ Mutagenesis Kit (Stratagene, according to manufacturer's instructions). Endogenous GPCR is preferably used as a template and two mutagenesis primers utilized, as well as, most preferably, a lysine mutagenesis oligonucleotide and a selection marker oligonucleotide (included in kit). For convenience, the codon mutation incorporated into the human GPCR and the respective oligonucleotides are noted, in standard

form (Table H):

TABLE H

Receptor Identifier	Codon Mutation	Lysine Mutagenesis (SEQ.ID.NO.)	5'-3' orientation, mutation underlined	Selection Marker (SEQ.ID.NO.)	5'-3' orientation
hCHN3	S284K	ATGGAGAAAAGAAATCAAAGAA	TGTTCTATATA (115)	TATATAGAACATTTCTTT	GAITCTTTCTCCAT
hCHN6	L352K	CGCTCTCTGGCCTTGAAGCGCAC	GCTCAGC (117)	GCTGAGCGTGGCTTCA	(116)
hCHN8	N235K	CCGAGAAAAGGTGAAAGTCA	AAGTTTC (119)	AGCCAGAGAGCG (118)	GAAACTTTGACTTTCAC
hCHN9	G223K	GGGCGCGGTGAAACGGCTGG	TOAGC (121)	CTTTTCTCTGGG (120)	GCTCACCAGCCGTTTCA
hCHN10	L231K	CCCTTGAAAAGCCTAAGAACTT	GGTCATC (123)	CCCGGCCCC (122)	GATGACCAAGTTCTTAG
					GCTTTTCAAGGGG (124)

Example 3

RECEPTOR EXPRESSION

Although a variety of cells are available to the art for the expression of proteins, it is most preferred that mammalian cells be utilized. The primary reason for this is predicated upon practicalities, i.e., utilization of, e.g., yeast cells for the expression of a GPCR, while possible, introduces into the protocol a non-mammalian cell which may not (indeed, in the case of yeast, does not) include the receptor-coupling, genetic-mechanism and secretory pathways that have evolved for mammalian systems - thus, results obtained in non-mammalian cells, while of potential use, are not as preferred as that obtained from mammalian cells. Of the mammalian cells, COS-7, 293 and 293T cells are particularly preferred, although the specific mammalian cell utilized can be predicated upon the particular needs of the artisan.

On day one, 1X10⁷ 293T cells per 150mm plate were plated out. On day two, two reaction tubes were prepared (the proportions to follow for each tube are per plate): tube A was prepared by mixing 20μg DNA (e.g., pCMV vector; pCMV vector with receptor cDNA, etc.) in 1.2ml serum free DMEM (Irvine Scientific, Irvine, CA); tube B was

prepared by mixing 120μl lipofectamine (Gibco BRL) in 1.2ml serum free DMEM. Tubes A and B were admixed by inversions (several times), followed by incubation at room temperature for 30-45min. The admixture is referred to as the "transfection mixture".

Plated 293T cells were washed with 1XPBS, followed by addition of 10ml serum free DMEM. 2.4ml of the transfection mixture were added to the cells, followed by incubation for 4hrs at 37°C/5% CO₂. The transfection mixture was removed by aspiration, followed by the addition of 25ml of DMEM/10% Fetal Bovine Serum. Cells were incubated at 37°C/5% CO₂. After 72hr incubation, cells were harvested and utilized for analysis.

Example 4

ASSAYS FOR DETERMINATION OF CONSTITUTIVE ACTIVITY OF NON-ENDOGENOUS GPCRS

A variety of approaches are available for assessment of constitutive activity of the non-endogenous human GPCRs. The following are illustrative; those of ordinary skill in the art are credited with the ability to determine those techniques that are preferentially beneficial for the needs of the artisan.

1. Membrane Binding Assays: [³⁵S]GTPγS Assay

When a G protein-coupled receptor is in its active state, either as a result of ligand binding or constitutive activation, the receptor couples to a G protein and stimulates release of GDP and subsequent binding of GTP to the G protein. The alpha subunit of the protein-receptor complex acts as a GTPase and slowly hydrolyzes the GTP to GDP, at which point the receptor normally is deactivated. Constitutively activated receptors continue to exchange GDP for GTP. The non-hydrolyzable GTP analog, [³⁵S]GTPγS, can be utilized to demonstrate enhanced binding of [³⁵S]GTPγS to membranes expressing constitutively activated receptors. The advantage of using [³⁵S]GTPγS binding to measure constitutive

- 43 -

activation is that: (a) it is generically applicable to all G protein-coupled receptors; (b) it is proximal at the membrane surface making it less likely to pick-up molecules which affect the intracellular cascade.

The assay utilizes the ability of G protein coupled receptors to stimulate [³⁵S]GTPγS binding to membranes expressing the relevant receptors. The assay can, therefore, be used in a direct identification method to screen candidate compounds to known, orphan and constitutively activated G protein-coupled receptors. The assay is generic and has application to drug discovery at all G protein-coupled receptors.

The [³⁵S]GTPγS assay can be incubated in 20 mM HEPES and between 1 and about 20mM MgCl₂ (this amount can be adjusted for optimization of results, although 20mM is preferred) pH 7.4, binding buffer with between about 0.3 and about 1.2 nM [³⁵S]GTPγS (this amount can be adjusted for optimization of results, although 1.2 is preferred) and 12.5 to 75 μg membrane protein (e.g. COS-7 cells expressing the receptor; this amount can be adjusted for optimization, although 75μg is preferred) and 1 μM GDP (this amount can be changed for optimization) for 1 hour. Wheatgerm agglutinin beads (25 μl; Amersham) should then be added and the mixture incubated for another 30 minutes at room temperature. The tubes are then centrifuged at 1500 x g for 5 minutes at room temperature and then counted in a scintillation counter.

A less costly but equally applicable alternative has been identified which also meets the needs of large scale screening. Flash plates™ and Wallac™ scintistrips may be utilized to format a high throughput [³⁵S]GTPγS binding assay. Furthermore, using this technique, the assay can be utilized for known GPCRs to simultaneously monitor tritiated ligand binding to the receptor at the same time as monitoring the efficacy via [³⁵S]GTPγS binding. This is

- 44 -

possible because the Wallac beta counter can switch energy windows to look at both tritium and ³⁵S-labeled probes. This assay may also be used to detect other types of membrane activation events resulting in receptor activation. For example, the assay may be used to monitor ³²P phosphorylation of a variety of receptors (both G protein coupled and tyrosine kinase receptors). When the membranes are centrifuged to the bottom of the well, the bound [³⁵S]GTPγS or the ³²P-phosphorylated receptor will activate the scintillant which is coated of the wells. Scint® strips (Wallac) have been used to demonstrate this principle. In addition, the assay also has utility for measuring ligand binding to receptors using radioactively labeled ligands. In a similar manner, when the radiolabeled bound ligand is centrifuged to the bottom of the well, the scintistrip label comes into proximity with the radiolabeled ligand resulting in activation and detection.

2. Adenylyl Cyclase

A Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) designed for cell-based assays can be modified for use with crude plasma membranes. The Flash Plate wells contain a scintillant coating which also contains a specific antibody recognizing cAMP. The cAMP generated in the wells was quantitated by a direct competition for binding of radioactive cAMP tracer to the cAMP antibody. The following serves as a brief protocol for the measurement of changes in cAMP levels in membranes that express the receptors.

Transfected cells are harvested approximately three days after transfection. Membranes were prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization is performed on ice using a Brinkman Polyturon™ for approximately 10 seconds. The resulting homogenate is centrifuged at 49,000

X g for 15 minutes at 4°C. The resulting pellet is then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of measurement, the membrane pellet is slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂ (these amounts can be optimized, although the values listed herein are preferred), to yield a final protein concentration of 0.60mg/ml (the resuspended membranes were placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 µCi of tracer [¹²⁵I] cAMP (100 µl) to 11 ml Detection Buffer) are prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer is prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 µM GTP (Sigma), and 0.2 mM ATP (Sigma); Assay Buffer can be stored on ice until utilized. The assay is initiated by addition of 50ul of assay buffer followed by addition of 50ul of membrane suspension to the NEN Flash Plate. The resultant assay mixture is incubated for 60 minutes at room temperature followed by addition of 100ul of detection buffer. Plates are then incubated an additional 2-4 hours followed by counting in a Wallac MicroBeta™ scintillation counter. Values of cAMP/well are extrapolated from a standard cAMP curve that is contained within each assay plate.

20 C. Reporter-Based Assays

1. CREB Reporter Assay (Gs-associated receptors)

A method to detect Gs stimulation depends on the known property of the transcription factor CREB, which is activated in a cAMP-dependent manner. A PathDetect™ CREB trans-

Reporting System (Stratagene, Catalogue # 219010) can utilized to assay for Gs coupled activity in 293 or 293T cells. Cells are transfected with the plasmids components of this above system and the indicated expression plasmid encoding endogenous or mutant receptor using a Mammalian Transfection Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 400 ng pFR-Luc (luciferase reporter plasmid containing Gal4 recognition sequences), 40 ng pFA2-CREB (Gal4-CREB fusion protein containing the Gal4 DNA-binding domain), 80 ng pCMV-receptor expression plasmid (comprising receptor) and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid; alkaline phosphatase activity is measured in the media of transfected cells to control for variations in transfection efficiency between samples) are combined in a calcium phosphate precipitate as per the Kit's instructions. Half of the precipitate is equally distributed over 3 wells in a 96-well plate, kept on the cells overnight, and replaced with fresh medium the following morning. Forty-eight (48) hr after the start of the transfection, cells are treated and assayed for, e.g., luciferase activity

15 2. AP1 reporter assay (Gq-associated receptors)

A method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing AP1 elements in their promoter. A Pathdetect™ AP-1 cis-Reporting System (Stratagene, Catalogue # 219073) can be utilized following the protocol set forth above with respect to the CREB reporter assay, except that the components of the calcium phosphate precipitate were 410 ng pAP1-Luc, 80 ng pCMV-receptor expression plasmid, and 20 ng CMV-SEAP.

3. CRE-Luc Reporter Assay

293 and 293T cells are plated-out on 96 well plates at a density of 2 x 10⁴ cells per

- 47 -

well and were transfected using Lipofectamine Reagent (BRL) the following day according to manufacturer instructions. A DNA/lipid mixture is prepared for each 6-well transfection as follows: 260ng of plasmid DNA in 100µl of DMEM were gently mixed with 2µl of lipid in 100µl of DMEM (the 260ng of plasmid DNA consisted of 200ng of a 8xCRE-Luc reporter plasmid (see below and Figure 1 for a representation of a portion of the plasmid), 50ng of CMV comprising endogenous receptor or non-endogenous receptor or pCMV alone, and 10ng of a GPRS expression plasmid (GPRS in pcDNA3 (Invitrogen)). The 8xCRE-Luc reporter plasmid was prepared as follows: vector SRF-β-gal was obtained by cloning the rat somatostatin promoter (-71/+51) at BglIV-HindIII site in the pβgal-Basic Vector (Clontech). Eight (8) copies of cAMP response element were obtained by PCR from an adenovirus template AdpCF126CCRE8 (see, 7 *Human Gene Therapy* 1883 (1996)) and cloned into the SRF-β-gal vector at the Kpn-BglIV site, resulting in the 8xCRE-β-gal reporter vector. The 8xCRE-Luc reporter plasmid was generated by replacing the beta-galactosidase gene in the 8xCRE-β-gal reporter vector with the luciferase gene obtained from the pGL3-basic vector (Promega) at the HindIII-BamHI site. Following 30 min. incubation at room temperature, the DNA/lipid mixture was diluted with 400 µl of DMEM and 100µl of the diluted mixture was added to each well. 100 µl of DMEM with 10% FCS were added to each well after a 4hr incubation in a cell culture incubator. The following day the transfected cells were changed to with 200 µl/well of DMEM with 10% FCS. Eight (8) hours later, the wells were changed to 100 µl/well of DMEM without phenol red, after one wash with PBS. Luciferase activity were measured the next day using the LucLite™ reporter gene assay kit (Packard) following manufacturer instructions and read on a 1450 MicroBeta™ scintillation and luminescence counter (Wallac).

- 48 -

4. SRF-LUC Reporter Assay

One method to detect Gq stimulation depends on the known property of Gq-dependent phospholipase C to cause the activation of genes containing serum response factors in their promoter. A Pathdetect™ SRF-Luc-Reporting System (Stratagene) can be utilized to assay for Gq coupled activity in, e.g., COS7 cells. Cells are transfected with the plasmid components of the system and the indicated expression plasmid encoding endogenous or non-endogenous GPCR using a Mammalian Transfection™ Kit (Stratagene, Catalogue #200285) according to the manufacturer's instructions. Briefly, 410 ng SRF-Luc, 80 ng pCMV-receptor expression plasmid and 20 ng CMV-SEAP (secreted alkaline phosphatase expression plasmid, alkaline phosphatase activity is measured in the media of transfected cells to control for variations in transfection efficiency between samples) are combined in a calcium phosphate precipitate as per the manufacturer's instructions. Half of the precipitate is equally distributed over 3 wells in a 96-well plate, kept on the cells in a serum free media for 24 hours. The last 5 hours the cells are incubated with 1µM Angiotensin, where indicated. Cells are then lysed and assayed for luciferase activity using a LucLite™ Kit (Packard, Cat. # 6016911) and "TriLux 1450 Microbeta" liquid scintillation and luminescence counter (Wallac) as per the manufacturer's instructions. The data can be analyzed using GraphPad Prism™ 2.0a (GraphPad Software Inc.).

5. Intracellular IP₃ Accumulation Assay

On day 1, cells comprising the receptors (endogenous and/or non-endogenous) can be plated onto 24 well plates, usually 1x10⁵ cells/well (although his number can be optimized. On day 2 cells can be transfected by firstly mixing 0.25ug DNA in 50 ul serum free DMEM/well and 2 ul lipofectamine in 50 µl serumfree DMEM/well. The solutions

are gently mixed and incubated for 15-30 min at room temperature. Cells are washed with 0.5 ml PBS and 400 μ l of serum free media is mixed with the transfection media and added to the cells. The cells are then incubated for 3-4 hrs at 37°C/5%CO₂ and then the transfection media is removed and replaced with 1ml/well of regular growth media. On day 3 the cells are labeled with ³H-myoinositol. Briefly, the media is removed and the cells are washed with 0.5 ml PBS. Then 0.5 ml inositol-free/serum free media (GIBCO BRL) is added/well with 0.25 μ Ci of ³H-myoinositol / well and the cells are incubated for 16-18 hrs o/n at 37°C/5%CO₂. On Day 4 the cells are washed with 0.5 ml PBS and 0.45 ml of assay medium is added containing inositol-free/serum free media 10 μ M pargyline 10 mM lithium chloride or 0.4 ml of assay medium and 50 μ l of 10x ketanserin (ket) to final concentration of 10 μ M. The cells are then incubated for 30 min at 37°C. The cells are then washed with 0.5 ml PBS and 200 μ l of fresh/icecold stop solution (1M KOH; 18 mM Na-borate; 3.8 mM EDTA) is added/well. The solution is kept on ice for 5-10 min or until cells were lysed and then neutralized by 200 μ l of fresh/ice cold neutralization sol. (7.5 % HCL). The lysate is then transferred into 1.5 ml eppendorf tubes and 1 ml of chloroform/methanol (1:2) is added/tube. The solution is vortexed for 15 sec and the upper phase is applied to a Biorad AG1-X8™ anion exchange resin (100-200 mesh). Firstly, the resin is washed with water at 1:1.25 W/V and 0.9 ml of upper phase is loaded onto the column. The column is washed with 10 mls of 5 mM myo-inositol and 10 ml of 5 mM Na-borate/60mM Na-formate. The inositol tris phosphates are eluted into scintillation vials containing 10 ml of scintillation cocktail with 2 ml of 0.1 M formic acid/ 1 M ammonium formate. The columns are regenerated by washing with 10 ml of 0.1 M formic acid/3M ammonium formate and rinsed twice with dd H₂O and stored at 4°C in water.

Exemplary results are presented below in Table I:

TABLE I

Receptor	Mutation	Assay Utilized	Signal Generated: Endogenous Version (Relative Light Units)	Signal Generated: Non-Endogenous Version (Relative Light Units)	Percent Difference
hAT1	F239K	SRF-LUC	34	137	75% ¹
	AT2K255C3	SRF-LUC	34	127	73% ¹
hTDAG8	I225K	CRE-LUC (293 cells)	2,715	14,440	81% ¹
	I225K	CRE-LUC (293T cells)	65,681	185,636	65% ¹
hH9	F236K	CRE-LUC	1,887	6,096	69% ¹
hCCKB	V332K	CRE-LUC	785	3,223	76% ¹

C. CELL-BASED DETECTION ASSAY (EXAMPLE -TDAG8)

293 cells were plated-out on 150mm plates at a density of 1.3×10^7 cells per plate, and were transfected using 12 μ g of the respective DNA and 60 μ l of Lipofectamine Reagent (BRL) per plate. The transfected cells were grown in media containing serum for an assay performed 24 hours post-transfection. For detection assay performed 48 hours transfection (assay comparing serum and serum-free media; see Figure 3), the initial media was changed to either serum or serum-free media. The serum-free media was comprised solely of Dulbecco's Modified Eagle's (DME) High Glucose Medium (Irvine Scientific #9024). In addition to the above DME Medium, the media with serum contained the following: 10% Fetal Bovine Serum (Hyclone #SH30071.03), 1% of 100mM Sodium Pyruvate (Irvine Scientific #9334), 1% of 20mM L-Glutamine (Irvine Scientific #9317), and 1% of Penicillin-

Streptomycin solution (Irvine Scientific #9366).

A 96-well Adenyl Cyclase Activation Flashplate™ was used (NEN: #SMP004A). First, 50ul of the standards for the assay were added to the plate, in duplicate, ranging from concentrations of 50pmol to zero pmol cAMP per well. The standard cAMP (NEN: #SMP004A) was reconstituted in water, and serial dilutions were made using 1xPBS (Irvine Scientific: #9240). Next, 50ul of the stimulation buffer (NEN: #SMP004A) was added to all wells. In the case of using compounds to measure activation or inactivation of cAMP, 10ul of each compound, diluted in water, was added to its respective well, in triplicate. Various final concentrations used range from 1uM up to 1mM. Adenosine 5'-triphosphate, ATP, (Research Biochemicals International: #A-141) and Adenosine 5'-diphosphate, ADP, (Sigma: #A2754) were used in the assay. Next, the 293 cells transfected with the respective cDNA (CMV or TDAG8) were harvested 24 (assay detection in serum media) or 48 hours post-transfection (assay detection comparing serum and serum-free media). The media was aspirated and the cells washed once with 1xPBS. Then 5ml of 1xPBS was added to the cells along with 3ml of cell dissociation buffer (Sigma: #C-1544). The detached cells were transferred to a centrifuge tube and centrifuged at room temperature for five minutes. The supernatant was removed and the cell pellet was resuspended in an appropriate amount of 1xPBS to obtain a final concentration of 2×10^6 cells per milliliter. To the wells containing the compound, 50ul of the cells in 1xPBS (1x10⁶ cells/well) were added. The plate was incubated on a shaker for 15 minutes at room temperature. The detection buffer containing the tracer cAMP was prepared. In 1ml of detection buffer (NEN: #SMP004A), 50ul (equal to 1uCi) of [¹²⁵I]cAMP (NEN: #SMP004A) was added. Following incubation, 50ul of this detection buffer containing tracer cAMP was added to each well. The plate was placed on a shaker and

incubated at room temperature for two hours. Finally, the solution from the wells of the plate were aspirated and the flashplate was counted using the Wallac MicroBeta™ scintillation counter.

In Figure 2A, ATP and ADP bind to endogenous TDAG8 resulting in an increase of cAMP of about 59% and about 55% respectively. Figure 2B evidences ATP and ADP binding to endogenous TDAG8 where endogenous TDAG8 was transfected and grown in serum and serum-free medium. ATP binding to endogenous TDAG8 grown in serum media evidences an increase in cAMP of about 65%, compared to the endogenous TDAG8 with no compounds; in serum-free media there was an increase of about 68%. ADP binding to endogenous TDAG8 in serum evidences about a 61% increase, while in serum-free ADP binding evidences an increase of about 62% increase. ATP and ADP bind to endogenous TDAG8 with an EC50 value of 139.8uM and 120.5uM, respectively (data not shown).

Although the results presented in Figure 2B indicate substantially the same results when serum and serum-free media were compared, our choice is to use a serum based media, although a serum-free media can also be utilized.

Example 6 GPCR FUSION PROTEIN PREPARATION

The design of the constitutively activated GPCR-G protein fusion construct was accomplished as follows: both the 5' and 3' ends of the rat G protein Gso (long form, 110kH. et al., 83 PNAS 3776 (1986)) were engineered to include a HindIII (5'-AAGCTT-3') sequence thereon. Following confirmation of the correct sequence (including the flanking HindIII sequences), the entire sequence was shuttled into pcDNA3.1(-) (Invitrogen, cat. no. V795-20) by subcloning using the HindIII restriction site of that vector. The correct

orientation for the Gsa sequence was determined after subcloning into pcDNA3.1(-). The modified pcDNA3.1(-) containing the rat Gsa gene at HindIII sequence was then verified; this vector was now available as a "universal" Gsa protein vector. The pcDNA3.1(-) vector contains a variety of well-known restriction sites upstream of the HindIII site, thus beneficially providing the ability to insert, upstream of the Gs protein, the coding sequence of an endogenous, constitutively active GPCR. This same approach can be utilized to create other "universal" G protein vectors, and, of course, other commercially available or proprietary vectors known to the artisan can be utilized - the important criteria is that the sequence for the GPCR be upstream and in-frame with that of the G protein.

10 TDAG8 couples via Gs, while H9 couples via Gz. For the following exemplary GPCR Fusion Proteins, fusion to Gsa was accomplished.

A TDAG8(I225K)-Gsa Fusion Protein construct was made as follows: primers were designed as follows:

5'-gatCTTAGAATGAACAGCACATGTATTGAAG-3' (SEQ.ID.NO.: 125; sense)

15 5'-cagGGTACCGCTCAAGGACCTCTAATTCCATAG-3' (SEQ.ID.NO.: 126; antisense).

Nucleotides in lower caps are included as spacers in the restriction sites between the G protein and TDAG8. The sense and anti-sense primers included the restriction sites for XbaI and KpnI, respectively.

PCR was then utilized to secure the respective receptor sequences for fusion within the Gsa universal vector disclosed above, using the following protocol for each: 100ng cDNA for TDAG8 was added to separate tubes containing 2ul of each primer (sense and anti-sense), 3uL of 10mM dNTPs, 10uL of 10XTaqPlus™ Precision buffer, 1uL of TaqPlus™ Precision polymerase (Stratagene: #600211), and 80uL of water. Reaction temperatures and cycle times for TDAG8 were as follows: the initial denaturing step was done at 94°C for five minutes, and

a cycle of 94°C for 30 seconds; 55°C for 30 seconds; 72°C for two minutes. A final extension time was done at 72°C for ten minutes. PCR product for was run on a 1% agarose gel and then purified (data not shown). The purified product was digested with XbaI and KpnI (New England Biolabs) and the desired inserts purified and ligated into the Gs universal vector at the respective restriction site. The positive clones was isolated following transformation and determined by restriction enzyme digest; expression using 293 cells was accomplished following the protocol set forth *infra*. Each positive clone for TDAG8:Gsa Fusion Protein was sequenced to verify correctness.

GPCR Fusion Proteins comprising non-endogenous, constitutively activated TDAG8(I225K) were analyzed as above and verified for constitutive activation.

An H9(F236K)-Gsa Fusion Protein construct was made as follows: primers were designed as follows:

5'-TTAgatcGGGGCCCAACCCCTAGCGGT-3' (SEQ.ID.NO.: 145; sense)

5'-ggtaacCCCCACAGCCATTTCATCAGGATC-3' (SEQ.ID.NO.: 146; antisense).

15 Nucleotides in lower caps are included as spacers in the restriction sites between the G protein and H9. The sense and anti-sense primers included the restriction sites for EcoRV and KpnI, respectively such that spacers (attributed to the restriction sites) exists between G protein and H9.

PCR was then utilized to secure the respective receptor sequences for fusion within the Gsa universal vector disclosed above, using the following protocol for each: 80ng cDNA for H9 was added to separate tubes containing 100ng of each primer (sense and anti-sense), and 45uL of PCR Supremix™ (Gibco-Brl, LifeTech) (50ul total reaction volume). Reaction temperatures and cycle times for H9 were as follows: the initial denaturing step was done at 94°C for one, and a cycle of 94°C for 30 seconds; 55°C for 30 seconds; 72°C for two

- 55 -

minutes. A final extension time was done at 72°C for seven minutes. PCR product for was run on a 1% agarose gel and then purified (data not shown). The purified product was cloned into pCRII-TOPO™ System followed by identification of positive clones. Positive clones were isolated, digested with EcoRV and KpnI (New England Biolabs) and the desired inserts were isolated, purified and ligated into the Gs universal vector at the respective restriction site.

The positive clones were isolated following transformation and determined by restriction enzyme digest, expression using 293 cells was accomplished following the protocol set forth *infra*. Each positive clone for H9(F236K)-Gs - Fusion Protein was sequenced to verify correctness. Membranes were frozen (-80°C) until utilized.

To ascertain the ability of measuring a cAMP response mediated by the Gs protein (even though H9 couples with G_z), the following cAMP membrane assay was utilized, based upon an NEN Adenylyl Cyclase Activation Flapplate™ Assay kit (96 well format). "Regeneration Buffer" consisted of 10mM HEPES, 100mM NaCl and 10mM MgCl (ph 7.4). "Regeneration Buffer" was prepared in Binding Buffer and consisted of 20mM phosphocreatine, 20U creatine phosphokinase, 20mM GTP, 0.2mM ATP, and 0.6mM IBMX. "cAMP Standards" were prepared in Binding Buffer as follows:

	cAMP Stock (5,000 pmol/ml in 2ml H ₂ O)	Added to indicated amount of Binding Buffer	Final Assay Concentration (50ul into 100ul)
	in ul	in ul	to achieve indicated pmol/well
20	250	500ul	50
A	500 of A	500ul	25
B	500 of B	500ul	12.5
C	500 of C	750ul	5.0
D	500 of D	500ul	2.5
E	500 of E	500ul	1.25
F	500 of F	750ul	0.5
25	500 of F	750ul	

Frozen membranes (both pCMV as control and the non-endogenous H(-Gs Fusion Protein) were thawed (on ice at room temperature until in solution). Membranes were

- 56 -

homogenized with a polytron until in suspension (2 x 15 seconds). Membrane protein concentration was determined using the Bradford Assay Protocol (*see infra*). Membrane concentration was diluted to 0.5mg/ml in Regeneration Buffer (final assay concentration - 25ug/well). Thereafter, 50ul of Binding Buffer was added to each well. For control, 50ul/well of cAMP standard was added to wells 11 and 12 A-G, with Binding Buffer alone to 12H (on the 96-well format). Thereafter, 50ul/well of protein was added to the wells and incubated at room temperature (on shaker) for 60min. 100ul [¹²⁵I]cAMP in Detection Buffer (*see infra*) was added to each well (final - 50ul [¹²⁵I]cAMP into 11ml Detection Buffer). These were incubated for 2hrs at room temperature. Plates were aspirated with an 8 channel manifold and sealed with plate covers. Results (pmoles cAMP bound) were read in a Wallac™ 1450 on "prot #15). Results are presented in Figure 3.

The results presented in Figure 3 indicate that the Gs coupled fusion was able to "drive" the cyclase reaction such that measurement of the constitutive activation of H9(F236K) was viable. Based upon these results, the direct identification of candidate compounds that are inverse agonists, agonists and partial agonists is possible using a cyclase-based assay.

Example 6 Protocol: Direct Identification of Inverse Agonists and Agonists Using [³⁵S]GTPγS

Although we have utilized endogenous, constitutively active GPCRs for the direct identification of candidate compounds as, *e.g.*, inverse agonists, for reasons that are not altogether understood, intra-assay variation can become exacerbated. Preferably, then, a GPCR Fusion Protein, as disclosed above, is also utilized with a non-endogenous, constitutively activated GPCR. We have determined that when such a protein is used, intra-assay variation appears to be substantially stabilized, whereby an effective signal-to-noise ratio is obtained. This has the beneficial result of allowing for a more robust identification

of candidate compounds. Thus, it is preferred that for direct identification, a GPCR Fusion Protein be used and that when utilized, the following assay protocols be utilized.

Membrane Preparation

Membranes comprising the non-endogenous, constitutively active orphan GPCR

5 Fusion Protein of interest and for use in the direct identification of candidate compounds as inverse agonists, agonists or partial agonists are preferably prepared as follows:

a. Materials

"Membrane Scrape Buffer" is comprised of 20mM HEPES and 10mM EDTA, pH 7.4;

"Membrane Wash Buffer" is comprised of 20 mM HEPES and 0.1 mM EDTA, pH 7.4;

10 "Binding Buffer" is comprised of 20mM HEPES, 100 mM NaCl, and 10 mM MgCl₂, pH 7.4

b. Procedure

All materials are kept on ice throughout the procedure. Firstly, the media is aspirated from a confluent monolayer of cells, followed by rinse with 10ml cold PBS, followed by aspiration. Thereafter, 5ml of Membrane Scrape Buffer is added to scrape cells; this is followed by transfer of cellular extract into 50ml centrifuge tubes (centrifuged at 20,000 rpm for 17 minutes at 4°C). Thereafter, the supernatant is aspirated and the pellet is resuspended in 30ml Membrane Wash Buffer followed by centrifuge at 20,000 rpm for 17 minutes at 4°C. The supernatant is then aspirated and the pellet resuspended in Binding Buffer. This is then homogenized using a Brinkman polytron™ homogenizer (15-20 second bursts until the all material is in suspension). This is referred to herein as "Membrane Protein".

Bradford Protein Assay

Following the homogenization, protein concentration of the membranes is determined using the Bradford Protein Assay (protein can be diluted to about 1.5mg/ml, aliquoted and

frozen (-80°C) for later use; when frozen, protocol for use is as follows: on the day of the assay, frozen Membrane Protein is thawed at room temperature, followed by vortex and then homogenized with a polytron at about 12 x 1,000 rpm for about 5-10 seconds; it is noted that for multiple preparations, the homogenizer should be thoroughly cleaned between homogenization of different preparations).

a. Materials

Binding Buffer (as per above); Bradford Dye Reagent; Bradford Protein Standard utilized, following manufacturer instructions (Biorad, cat. no. 500-0006).

b. Procedure

10 Duplicate tubes are prepared, one including the membrane, and one as a control "blank". Each contained 800ul Binding Buffer. Thereafter, 10ul of Bradford Protein Standard (1mg/ml) is added to each tube, and 10ul of membrane Protein is then added to just one tube (not the blank). Thereafter, 200ul of Bradford Dye Reagent is added to each tube, followed by vortex of each. After five (5) minutes, the tubes were re-vortexed and the material therein is transferred to cuvettes. The cuvettes are then read using a CECIL 3041 spectrophotometer, at wavelength 595.

Direct Identification Assay

a. Materials

GDP Buffer consists of 37.5 ml Binding Buffer and 2mg GDP (Sigma, cat. no. G-7127), followed by a series of dilutions in Binding Buffer to obtain 0.2 uM GDP (final concentration of GDP in each well was 0.1 uM GDP); each well comprising a candidate compound, has a final volume of 200ul consisting of 100ul GDP Buffer (final concentration, 0.1uM GDP), 50ul Membrane Protein in Binding Buffer, and 50ul [³⁵S]GTPγS (0.6 nM) in

- 59 -

Binding Buffer (2.5 μ l [35 S]GTP γ S per 10ml Binding Buffer).

b. Procedure

Candidate compounds are preferably screened using a 96-well plate format (these can be frozen at -80°C). Membrane Protein (or membranes with expression vector excluding the

GPCR Fusion Protein, as control), are homogenized briefly until in suspension. Membrane concentration is then determined using the Bradford Protein Assay set forth above. Membrane Protein (and control) is then diluted to 0.25mg/ml in Binding Buffer (final assay concentration, 12.5 μ g/well). Thereafter, 100 μ l GDP Buffer is added to each well of a Wallac Scintistrip™ (Wallac). A 5 μ l pin-tool is then used to transfer 5 μ l of a candidate compound into such well (i.e., 5 μ l in total assay volume of 200 μ l is a 1:40 ratio such that the final screening concentration of the candidate compound is 10nM). Again, to avoid contamination, after each transfer step the pin tool should be rinsed in three reservoirs comprising water (1X), ethanol (1X) and water (2X) - excess liquid should be shaken from the tool after each rinse and dried with paper and kimwipes. Thereafter, 50 μ l of Membrane Protein is added to each well (a control well comprising membranes without the GPCR Fusion Protein is also utilized), and pre-incubated for 5-10 minutes at room temperature. Thereafter, 50 μ l of [35 S]GTP γ S (0.6 nM) in Binding Buffer is added to each well, followed by incubation on a shaker for 60 minutes at room temperature (again, in this example, plates were covered with foil). The assay is then stopped by spinning of the plates at 4000 RPM for 15 minutes at 22°C. The plates are then aspirated with an 8 channel manifold and sealed with plate covers. The plates are then read on a Wallac 1450 using setting "Prot. #37" (as per manufacturer instructions).

Example 7 Protocol: Confirmation Assay

Using an independent assay approach to provide confirmation of a directly identified

- 60 -

candidate compound as set forth above, it is preferred that a confirmation assay then be utilized. In this case, the preferred confirmation assay is a cyclase-based assay.

A modified Flash Plate™ Adenylyl Cyclase kit (New England Nuclear; Cat. No. SMP004A) is preferably utilized for confirmation of candidate compounds directly identified as inverse agonists and agonists to non-endogenous, constitutively activated orphan GPCRs in accordance with the following protocol.

Transfected cells are harvested approximately three days after transfection. Membranes are prepared by homogenization of suspended cells in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂. Homogenization is performed on ice using a Brinkman Polytron™ for approximately 10 seconds. The resulting homogenate is centrifuged at 49,000 X g for 15 minutes at 4°C. The resulting pellet is then resuspended in buffer containing 20mM HEPES, pH 7.4 and 0.1 mM EDTA, homogenized for 10 seconds, followed by centrifugation at 49,000 X g for 15 minutes at 4°C. The resulting pellet can be stored at -80°C until utilized. On the day of direct identification screening, the membrane pellet is slowly thawed at room temperature, resuspended in buffer containing 20mM HEPES, pH 7.4 and 10mM MgCl₂, to yield a final protein concentration of 0.60mg/ml (the resuspended membranes are placed on ice until use).

cAMP standards and Detection Buffer (comprising 2 μ Ci of tracer [35] cAMP (100 μ l) to 11 ml Detection Buffer) are prepared and maintained in accordance with the manufacturer's instructions. Assay Buffer is prepared fresh for screening and contained 20mM HEPES, pH 7.4, 10mM MgCl₂, 20mM phosphocreatine (Sigma), 0.1 units/ml creatine phosphokinase (Sigma), 50 μ M GTP (Sigma), and 0.2 mM ATP (Sigma). Assay Buffer can be stored on ice until utilized.

Candidate compounds identified as per above (if frozen, thawed at room temperature) are added, preferably, to 96-well plate wells (3 μ l/well; 12 μ M final assay concentration), together with 40 μ l Membrane Protein (30 μ g/well) and 50 μ l of Assay Buffer. This admixture is then incubated for 30 minutes at room temperature, with gentle shaking.

Following the incubation, 100 μ l of Detection Buffer is added to each well, followed by incubation for 2-24 hours. Plates are then counted in a Wallac MicroBeta™ plate reader using "Prot. #31" (as per manufacturer instructions).

It is intended that each of the patents, applications, and printed publications mentioned in this patent document be hereby incorporated by reference in their entirety.

As those skilled in the art will appreciate, numerous changes and modifications may be made to the preferred embodiments of the invention without departing from the spirit of the invention. It is intended that all such variations fall within the scope of the invention.

Although a variety of expression vectors are available to those in the art, for

purposes of utilization for both the endogenous and non-endogenous human GPCRs, it is most preferred that the vector utilized be pCMV. This vector was deposited with the American Type Culture Collection (ATCC) on October 13, 1998 (10801 University Blvd., Manassas, VA 20110-2209 USA) under the provisions of the Budapest Treaty for the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure. The DNA was tested by the ATCC and determined to be. The ATCC has assigned the following deposit number to pCMV: ATCC #203351.

CLAIMS

What is claimed is:

1. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hARE-3(F313K).
2. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 1.
3. A Plasmid comprising a Vector and the cDNA of claim 1.
4. A Host Cell comprising the Plasmid of claim 3.
5. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hARE-4(V233K).
6. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 5.
7. A Plasmid comprising a Vector and the cDNA of claim 5.
8. A Host Cell comprising the Plasmid of claim 7.
9. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hARE-5(A240K).
10. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 9.
11. A Plasmid comprising a Vector and the cDNA of claim 5.
12. A Host Cell comprising the Plasmid of claim 11.
13. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hGPCR14(L257K).

- 63 -

14. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 13.
15. A Plasmid comprising a Vector and the cDNA of claim 13.
16. A Host Cell comprising the Plasmid of claim 15.
17. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hGPCR27(C283K).
18. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 17.
19. A Plasmid comprising a Vector and the cDNA of claim 17.
20. A Host Cell comprising the Plasmid of claim 19.
21. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hARE-1(E232K).
22. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 21.
23. A Plasmid comprising a Vector and the cDNA of claim 21.
24. A Host Cell comprising the Plasmid of claim 23.
25. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hARE-2(G285K).
26. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 25.
27. A Plasmid comprising a Vector and the cDNA of claim 25.
28. A Host Cell comprising the Plasmid of claim 27.

- 64 -

29. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hPPRI(L239K).
30. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 29.
31. A Plasmid comprising a Vector and the cDNA of claim 29.
32. A Host Cell comprising the Plasmid of claim 31.
33. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hG2A(K232A).
34. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 33.
35. A Plasmid comprising a Vector and the cDNA of claim 33.
36. A Host Cell comprising the Plasmid of claim 35.
37. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hRUP3(L224K).
38. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 37.
39. A Plasmid comprising a Vector and the cDNA of claim 37.
40. A Host Cell comprising the Plasmid of claim 39.
41. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hRUP5(A236K).
42. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 41.
43. A Plasmid comprising a Vector and the cDNA of claim 41.

44. A Host Cell comprising the Plasmid of claim 42.
45. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hRUP6(N267K)
46. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 45.
47. A Plasmid comprising a Vector and the cDNA of claim 45.
48. A Host Cell comprising the Plasmid of claim 47.
49. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hRUP7(A302K).
50. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 49.
51. A Plasmid comprising a Vector and the cDNA of claim 49.
52. A Host Cell comprising the Plasmid of claim 51.
53. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hCHN4(V236K).
54. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 53.
55. A Plasmid comprising a Vector and the cDNA of claim 53.
56. A Host Cell comprising the Plasmid of claim 55.
57. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hMC4(A244K).
58. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 57.

59. A Plasmid comprising a Vector and the cDNA of claim 57.
60. A Host Cell comprising the Plasmid of claim 60.
61. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hCHN3(S284K).
62. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 61.
63. A Plasmid comprising a Vector and the cDNA of claim 61.
64. A Host Cell comprising the Plasmid of claim 63.
65. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hCHN6(L352K).
66. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 65.
67. A Plasmid comprising a Vector and the cDNA of claim 65.
68. A Host Cell comprising the Plasmid of claim 67.
69. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hCHN8(N235K).
70. A non-endogenous version of a human G protein-coupled receptor encoded by the cDNA of claim 69.
71. A Plasmid comprising a Vector and the cDNA of claim 69.
72. A Host Cell comprising the Plasmid of claim 71.
73. A cDNA encoding a non-endogenous, constitutively activated version of a human G protein-coupled receptor comprising hH9(F236K).
74. A non-endogenous version of a human G protein-coupled receptor encoded by the

- 67 -

cDNA of claim 73.

75. A Plasmid comprising a Vector and the cDNA of claim 73.

76. A Host Cell comprising the Plasmid of claim 74.

77. A cDNA encoding a non-endogenous, constitutively activated version of a human

G protein-coupled AT₁ receptor selected from the group consisting of:

hAT1(F239K); hAT1(N111A); hAT1(AT2K251C3); and hAT1(A243+).

78. A non-endogenous version of a human G protein-coupled receptor encoded by a

cDNA of claim 77.

79. A Plasmid comprising a Vector and the cDNA of claim 77.

80. A Host Cell comprising the Plasmid of claim 79.

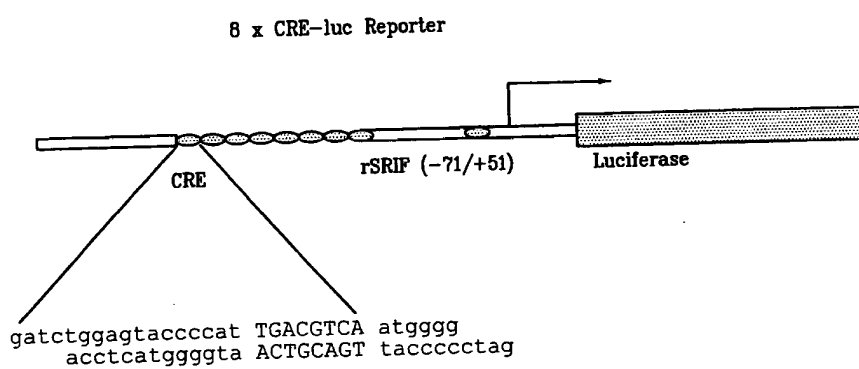
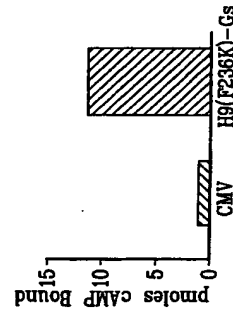
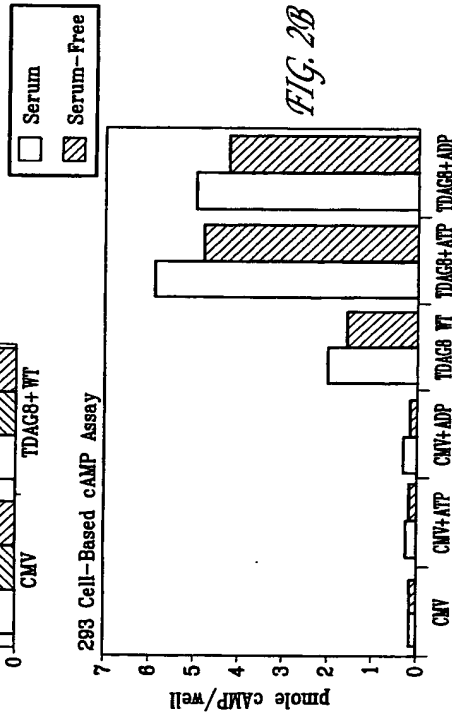
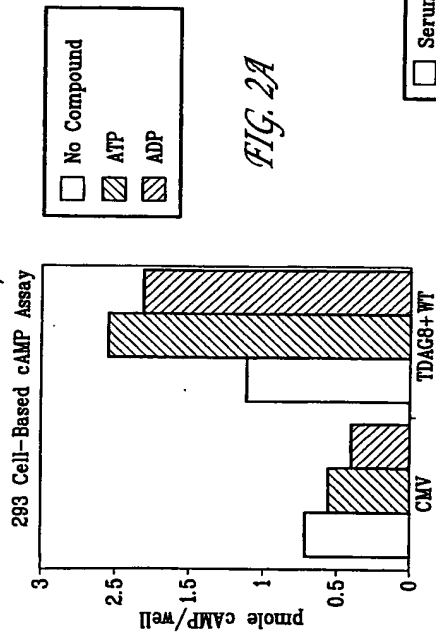


FIG. 1

2/2



SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Behan, Dominic P.
Lehmann-Bruinsma, Karin
Chalmers, Derek T.
Lowitz, Kevin P.
Lin, I-Lin
Dang, Ruong T.
Chen, Ruoping
Liaaw, Chen W.
Gore, Martin J.
White, Carol

- (ii) TITLE OF INVENTION: Non-Endogenous, Constitutively Activated Human G Protein-Coupled Receptors

- (iii) NUMBER OF SEQUENCES: 146

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: Arena Pharmaceuticals, Inc.
(B) STREET: 6166 Nancy Ridge Drive
(C) CITY: San Diego
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 92121

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: US
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: Burgoon, Richard P.
(B) REGISTRATION NUMBER: 34,787

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (858) 453-7200
(B) TELEFAX: (858) 453-7210

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1260 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

-2-

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGGCTCTTCT GGGGAGTGT GACTGCGTTC CATACCGGGA CATCCACAC AACATTGTC 60
 5 GTGTATGAAA AACACTACAT GAATTTACA CTCCTCCAC CATCCAGCA TCCGACCTC 120
 AGTCACATTC TTAGATATAG TTTTGAAAAC ATGGCTCCCA CTGGTTTGA TTTCTTGACC 180
 GTGAAATAGTA GAGCTGTGCC CACAACACA GCGACATTTA AGAGCTTAAA CTTCCTCTTT 240
 CAGATCACCC TTCTGCTAT AATGATATTC ATTCTGTTG TGTCTTTCT TGGGAACTTG 300
 GTTGTGGCC TCATGTTTA CCAAAAAGCT GCCATGAGT CTGCATTTA GATCTCTCTT 360
 10 GCCAGCCTAG CTTTGCAGA CATGTGCTT GCAAGTCTGA ACATGCCCTT TGCCTTGSTA 420
 ACTATCTTA CTACCCGATG GATTTTGGG AAATCTCTT GTAGGGTATC TGTATGTTT 480
 TTCTGTTAT TTGTGATAGA AGGATGAGC ATCTGCTCA TCATTAGAT AGATAGTTT 540
 CTATATATAG TCCAGAGGA GGAATAGCTA AACCCATATA GACTTAAGT TCTGATGCA 600
 GTTCTTGGG CAATCTCTT TTGTGAGCT TTTCTTTAG CCGTAGGAAA CCCCAGCTG 660
 15 CAGATACCTT CCGGAGCTCC CCAGTGTGT TTTGGGTACA CAACCAATC AGGCTACGAG 720
 GCTATGTTGA TTTTGATTTT TCTCATTTCT TTCTCAATAC CTTTCCGTGT AATATGTATC 780
 TCATTATAG GCAATCTCA CACCTTGG GACAATGCTT TGAAGATCCA TAGTAACCT 840
 GAAGTATAT GCTCTAGCA GGCAGGAAA CTGGGTCTCA TGAATCTGA GAGACCTTC 900
 CAGATGAGCA TTGACATGGG CTTTAAACA CGTGCCTCA CCACATTTT GATTCCTTT 960
 GCTGCTTCA TTGTCTGCTG GGGCCCATC AACCACTACA GCTTGTGGC AACATCAGT 1020
 AAGCACTTT ACTATACACA CAATTTTTT GAGATGACA CCGTGTACT GTGGCTCTGC 1080
 TACTCAAGT CTGCATGAA TCCGCTGATC TACTACTGGA GGAATTAAGA ATTCCATGAT 1140
 GCTTGCTGG ACATGATGCC TAGTCTTC AGATTGTC CCGAGCTCC TGTCTACACA 1200
 AACGACGGA TAGCTCTAG TGGTGTCTAT GTGTGTGGG AACATCGAC GGTGTGTGA 1260

25 (3) INFORMATION FOR SEQ ID NO:2:
 (A) LENGTH: 419 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

30

-3-

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Val Phe Ser Ala Val Leu Thr Ala Phe His Thr Gly Thr Ser Asn 1
 1 5 10 15
 5 Thr Thr Phe Val Val Tyr Glu Asn Thr Tyr Met Asn Ile Thr Leu Pro 20
 25
 Pro Pro Phe Glu His Pro Asp Leu Ser Pro Leu Leu Arg Tyr Ser Phe 35
 40
 10 Glu Thr Met Ala Pro Thr Gly Leu Ser Ser Leu Thr Val Asn Ser Thr 50
 55
 Ala Val Pro Thr Thr Pro Ala Ala Phe Lys Ser Leu Asn Leu Pro Leu 65
 70 75 80
 Glu Ile Thr Leu Ser Ala Ile Met Ile Phe Ile Leu Phe Val Ser Phe 85
 90 95
 15 Leu Gly Asn Leu Val Val Cys Leu Met Val Tyr Glu Lys Ala Ala Met 100
 105
 Arg Ser Ala Ile Asn Ile Leu Leu Ala Ser Leu Ala Phe Ala Asp Met 115
 120 125
 20 Leu Leu Ala Val Leu Asn Met Pro Phe Ala Leu Val Thr Ile Leu Thr 130
 135 140
 Thr Arg Tyr Ile Phe Gly Lys Phe Phe Cys Arg Val Ser Ala Met Phe 145
 150 155 160
 Phe Trp Leu Phe Val Ile Glu Gly Val Ala Ile Leu Leu Ile Ile Ser 165
 170 175
 25 Ile Asp Arg Phe Leu Ile Ile Val Glu Arg Glu Asp Lys Leu Asn Pro 180
 185 190
 Tyr Arg Ala Lys Val Leu Ile Ala Val Ser Trp Ala Thr Ser Phe Cys 195
 200 205
 30 Val Ala Phe Pro Leu Ala Val Gly Asn Pro Asp Leu Glu Ile Pro Ser 210
 215 220
 Arg Ala Pro Glu Cys Val Phe Gly Tyr Thr Thr Asn Pro Gly Tyr Glu 225
 230 235 240
 Ala Tyr Val Ile Leu Ile Ser Leu Ile Ser Phe Ile Pro Phe Leu 245
 250 255
 35 Val Ile Leu Tyr Ser Phe Met Gly Ile Leu Asn Thr Leu Arg His Asn 260
 265 270

- 4 -

Ala Leu Arg Ile His Ser Tyr Pro Glu Gly Ile Cys Leu Ser Gln Ala
275 280 285

Ser Lys Leu Gly Leu Met Ser Leu Gln Arg Pro Phe Gln Met Ser Ile
290 295 300

Asp Met Gly Phe Lys Thr Arg Ala Phe Thr Thr Ile Leu Ile Leu Phe
305 310 315 320

Ala Val Phe Ile Val Cys Trp Ala Pro Phe Thr Thr Tyr Ser Leu Val
325 330 335

Ala Thr Phe Ser Lys His Phe Tyr Tyr Gln His Asn Phe Phe Glu Ile
340 345 350

Ser Thr Trp Leu Leu Trp Leu Cys Tyr Leu Lys Ser Ala Leu Asn Pro
355 360 365

Leu Ile Tyr Tyr Trp Arg Ile Lys Lys Phe His Asp Ala Cys Leu Asp
370 375 380

Met Met Pro Lys Ser Phe Lys Phe Leu Pro Gln Leu Pro Gly His Thr
385 390 395 400

Lys Arg Arg Ile Arg Pro Ser Ala Val Tyr Val Cys Gly Glu His Arg
405 410 415

Thr Val Val

(4) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 119 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGTTAGCCA ACAGTCTCTC AACCAACAGT TCTGTTCTCC CGTGCTCTGA CTACCGACCT 60

30 ACCCACCACC TGCACTTGCT GGTCTACAGC TTGGTCTCTG CTGCCGGGCT CCCCCTCAAC 120

GCCTAGACCC TCTGGGTCTT CTTGGCGGCG CTGGCGGTGC ACTCGGTGGT GAGCGGTAC 180

ATGTGTAACC TGGGGGCCAG CGACCTGCTC TTCAACCTCT CGCTGCCCGT TCGTCTCTCC 240

TACTAGGCAC TGACACACTG GCCCTTCCCC GACCTCTGT GCCAGACGAC GGGCGCCATC 300

TTCCAGATGA ACATGTACGG CAGTGCATC TTCTGATGC TCATCAAGT GGNACGCTAC 360

- 5 -

GCGGCCATCG TGCACCCGCT GCGACTGGCG CACCTGCGGC GGGCCCGGCTG 420

CTCTGCTGG GCGTGTGGG GCTCATCTG GTGTTGCGG TGCCCGCCGC CGCGTGCAC 480

AGGCCCTGCG GTTGGCGCTA CCGGACCTC GAGGTGCGCC TATGCTTCTGA GAGCTTCAGC 540

GACGAGCTGT GGAAAGGACG GCTGCTGCC CTCGTGCTGC TGGCCGAGGC GCTGGGCTTC 600

5 CTGCTGCCCC TGGCGGCGGT GGTCTACTCG TCGGCGCGAG TCTTCTGGAC GCTGCGCGGC 660

CCCGACGCCA CCGAGAGCCA GCGCGCGCG AAGACCTGTC GCCTCTCTGCT GGTCAACCTC 720

GTCTCTTCCC TGCTGTGCTT CGTGCCCTAC AACAGACGCG TGGCGGTCTA CCGGCTGCTG 780

CGGAGCAAGC TGGTGGCGGC CAGGTGCTCT GCCCGCATC GCGTGGCGCG GGTGCTCATG 840

GTGATGTGTC TGCTGGCGCG CGCCAACTGC GTGCTGAGAC CGCTGGTGTGA CTACTTTAGC 900

10 GCCGAGGGCT TCCGCAACAC CTTGCGGCG CTGGGCACTC CGCACCGGGC CAGGACCTCG 960

GCCACCAACG GAGACGCGGC GCGGCTCGCG CAATCCGAAA GTTCCGCCGT CACCACCGAC 1020

GCCACCAACG CGATGCGCG CAGTCAAGGG CTGCTCCGAC CTTCCGACTC CCACCTCTCG 1080

TCCTCTCTCA CACAGTGTC CCAGGATCCC GCCCTCTGA 1119

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 372 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Leu Ala Asn Ser Ser Thr Asn Ser Ser Val Leu Pro Cys Pro 15
1 5 10

Asp Tyr Arg Pro Thr His Arg Leu His Leu Val Val Tyr Ser Leu Val 30
20 25

Leu Ala Ala Gly Leu Pro Leu Asn Ala Leu Ala Leu Trp Val Phe Leu 45
35 40

Arg Ala Leu Arg Val His Ser Val Val Ser Val Tyr Met Cys Asn Leu 60
50 55

Ala Ala Ser Asp Leu Leu Phe Thr Leu Ser Leu Pro Val Arg Leu Ser 80
65 70 75

Tyr Tyr Ala Leu His His Tyr Pro Phe Pro Asp Leu Leu Cys Gln Thr

- 6 -

85 90 95

Thr Gly Ala Ile Phe Gln Met Asn Met Tyr Gly Ser Cys Ile Phe Leu
100 105 110

Met Leu Ile Asn Val Asp Arg Tyr Ala Ala Ile Val His Pro Leu Arg
115 120 125

Leu Arg His Leu Arg Arg Pro Arg Val Ala Arg Leu Leu Cys Leu Gly
130 135 140

Val Trp Ala Leu Ile Leu Val Phe Ala Val Pro Ala Ala Arg Val His
145 150 155 160

Arg Pro Ser Arg Cys Arg Tyr Arg Asp Leu Gln Val Arg Leu Cys Phe
165 170 175

Glu Ser Phe Ser Asp Gln Leu Trp Lys Gly Arg Leu Leu Pro Leu Val
180 185 190

Leu Leu Ala Gln Ala Leu Gly Phe Leu Leu Pro Leu Ala Ala Val Val
195 200 205

Tyr Ser Ser Gly Arg Val Phe Trp Thr Leu Ala Arg Pro Asp Ala Thr
210 215 220

Gln Ser Gln Arg Arg Lys Thr Val Arg Leu Leu Leu Ala Asn Leu
225 230 235 240

Val Ile Phe Leu Leu Cys Phe Val Pro Tyr Asn Ser Thr Leu Ala Val
245 250 255

Tyr Gly Leu Leu Arg Ser Lys Leu Val Ala Ser Val Pro Ala Arg
260 265 270

Asp Arg Val Arg Gly Val Leu Met Val Met Val Leu Leu Ala Gly Ala
275 280 285

Asn Cys Val Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ala Gln Gly Phe
290 295 300

Arg Asn Thr Leu Arg Gly Leu Gly Thr Pro His Arg Ala Arg Thr Ser
305 310 315 320

Ala Thr Asn Gly Thr Arg Ala Ala Leu Ala Gln Ser Gln Arg Ser Ala
325 330 335

Val Thr Thr Asp Ala Thr Arg Pro Asp Ala Ala Ser Gln Gly Leu Leu
340 345 350

Arg Pro Ser Asp Ser His Ser Leu Ser Ser Phe Thr Gln Cys Pro Gln
355 360 365

Asp Ser Ala Leu
370

- 7 -

(6) INFORMATION FOR SEQ ID NO:5:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1107 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATGGCCAACT CCACAGGAGT GAACGCTCA GAAGTCGAG GTCGTGTGG GTTATCTTG 60

GCAGCTGTG TGAGAGTGG GGCACTGCTG GGCAAGGAG CGTGTGTGT GTGTGTCTG 120

CGCAGCGCG GACTGCGGA CGCGCTTAC CTGCGGCACT TGTGAGTGT GAACCTGCTG 180

CGGCGCGCT CCATCATGCC GCTGAGGCTTG CTGCGGCACT CGCGGCGCG GCTGAGGCGG 240

GTCGCGCTTG GCCCGGAGC ATGCGGCGC GCTGAGTTC TCTCGGCGG TCTGCTGCG 300

GCCTGACAG TGGGGGTGAC CGCACTTGG CTGCGAGCT ACCGCTTCAT CGTGCACCG 360

CTGCGGCGAG GCTCGGAGC GCGGCTGTG CTCGTGCTCA CGCGGTGTG GCGCGGCGG 420

GGACTGTGG GCGGCTTTC CTTGCTGAG CCGCGGCGG CACGCGCGG TGTCTTGTCT 480

CGCTGTGCTG TCTTGGCTTG GGGGCTTGG CCTTCCGAG CGCTGTGAG CCTGTGAGC 540

TTGCGGCTGC CCGGCTTCT GCTGCTGAG GCTTACGAG GCATTTTGT GGTGCGGCT 600

CGGCTGAGC TGAAGCCCC ACAGCGGAG CAGGAGTCC GACTCGGCT GGACTCTCTG 660

GATGAGCGCC TTTCATCTT GCGGCGGCTC CGGCTGAGC TGCGGAGGG CAAGCGGCG 720

CTGCGGCGAG CGCTGAGCGT GGGCAATTG GCAAGCTGCT GAGTGCCTTA TGGTGCAGG 780

TGCTTGGAG CCGAGAGCG GCGCGGAGA GCCGAAGCG CTGTCACTTG GTTCGCTTAC 840

TGCGCTTGG CGGCTACCC CTTCCTGTAC GGGCTGCTGC AGGCGCGGT GCGCTTGGCA 900

CTGCGGCGCG TCTGTGCGG TGCACTGCTT GAACTGTGAC GAGGCTGAC TCCGCAAGC 960

TGGCAGCGCG GAGCACTTT GCAATGCTTC CAGAGACCC CAGAGGCGC TGCGGTAGGC 1020

CCTTGTAGG CTCGAGACA GACCCCGAG TTGCGAGAG GAGGAGGCC CGATACAG 1080

GGGCGACCTG AGAGTTCTT CTCTTGA
1107

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 368 amino acids

- 8 -

- (B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Ala Asn Ser Thr Gly Leu Asn Ala Ser Glu Val Ala Gly Ser Leu 15
1 5 10
Gly Leu Ile Leu Ala Ala Val Val Glu Val Gly Ala Leu Leu Gly Asn 30
20 25
Gly Ala Leu Leu Val Val Leu Arg Thr Pro Gly Leu Arg Asp Ala 45
35 40
Leu Tyr Leu Ala His Leu Cys Val Val Asp Leu Leu Ala Ala Ser 60
50 55
Ile Met Pro Leu Gly Leu Leu Ala Ala Pro Pro Gly Leu Gly Arg 80
65 70 75
Val Arg Leu Gly Pro Ala Pro Cys Arg Ala Ala Arg Phe Leu Ser Ala 95
85 90
Ala Leu Leu Pro Ala Cys Thr Leu Gly Val Ala Ala Leu Gly Leu Ala 110
100 105
Arg Tyr Arg Leu Ile Val His Pro Leu Arg Pro Gly Ser Arg Pro Pro 125
115 120
Pro Val Leu Val Leu Thr Ala Val Trp Ala Ala Gly Leu Leu Gly 140
130 135
Ala Leu Ser Leu Leu Gly Pro Pro Ala Pro Pro Pro Ala Pro Ala 160
145 150 155
Arg Cys Ser Val Leu Ala Gly Gly Leu Gly Pro Phe Arg Pro Leu Trp 175
165 170
Ala Leu Leu Ala Phe Ala Leu Pro Ala Leu Leu Leu Gly Ala Tyr 190
180 185
Gly Gly Ile Phe Val Val Ala Arg Arg Ala Ala Leu Arg Pro Pro Arg 205
195 200
Pro Ala Arg Gly Ser Arg Leu Arg Ser Asp Ser Leu Asp Ser Arg Leu 220
210 215
Ser Ile Leu Pro Pro Leu Arg Pro Arg Leu Pro Gly Gly Lys Ala Ala 240
225 230 235
Leu Ala Pro Ala Leu Ala Val Gly Gln Phe Ala Ala Cys Trp Leu Pro

- 9 -

Tyr Gly Cys Ala Cys Leu Ala Pro Ala Ala Arg Ala Ala Glu Ala Glu 270
260 265 270 275

Ala Ala Val Thr Trp Val Ala Tyr Ser Ala Phe Ala Ala His Pro Phe 285
275 280
Leu Tyr Gly Leu Leu Gln Arg Pro Val Arg Leu Ala Leu Gly Arg Leu 300
290 295 300

Ser Arg Arg Ala Leu Pro Gly Pro Val Arg Ala Cys Thr Pro Gln Ala 320
305 310 315
Trp His Pro Arg Ala Leu Leu Gln Cys Leu Gln Arg Pro Pro Glu G 335
325 330

Pro Ala Val Gly Pro Ser Glu Ala Pro Glu Gln Thr Pro Glu Leu Ala 350
340 345

Gly Gly Arg Ser Pro Ala Tyr Gln Gly Pro Pro Glu Ser Ser Leu Ser 365
355 360

(8) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1008 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATGGAATCAT CTTTCTCAT TGAAGTATC CTGTGTGTCC TGGCCTCCCT CATCATGCT 60
25 ACTACACAC TAGTGGCTGT GGCTGTGCTG CTGTTGATCC ACAAGATGA TGGTGTCACT 70
CTCTGCTTCA CTTGAATCT GGCTGTGGCT GACACCTTGA TTGGTGTGGC CATCTCTGGC 80
CTACTACAG ACCAGCTCTC CAGCCCTTCT CGGCCACAC AGAAGACCTT GTGACGCTG 240
CGAGTGGCAT TTGTCACTTC CTCGCAGCT GCCTGTGTCC TCACGGTCAT GCTGATCACC 300
TTTGACAGT ACCTTGCCAT CAGCAGGCC TTCCGCTACT TGAAGATCAT GAGTGGGTTT 360
GTGGCCGGGG CTTGCATTGC CGGCTGTGG TTAGTGTCTT ACCTCATTTG CTTCTCCCA 420
CTCGAATCC CCATGTTCCA GCAGACTGCC TACAAGGGC AGTGCAGCTT CTTTGTCTGA 480
TTTACGCTTC ACTTGTGTCT GACCTCTCC TGGTGTGGCT TCTTCCAGC CATGCTCCTC 540
TTTGTCTTCT TCTACTGCGA CATGCTCAAG ATTGCCTCCA TGCACAGCCA GCAGATTCCA 600

- 10 -

AAGATGAGAC ATGCAGAGC CATTGCTGGA GGTATTCAT CCCACAGGAC TCCACGAC 660
 TTCAAAGTC TCCGACTGCT GTCTGTCTC ATTGGAGCT TTGCTTATC CTGGACCC 720
 TTCTTATCA CTGGCAATGT GCAGATGACC TGCCAGAGT GTCACTCTTA CCTAGTCTG 780
 GAAAGTACC TGTGCTGCT CGAGTGGGC AACTCCCTGC TCAACCACT CATCTATGCC 840
 5 TATTGGACA AGAGATGCG ACTGCAGCTC TACCACATGG CCTTAGAGT GAAGAAGTG 900
 CTCACCTCAT TCTCTCTT TCTCTGGCC AGAATTTGG GCCCAGAGAG GCCCAGGAAA 960
 AGTTCCTGTC ACATGCTCAC TATCTCCAGC TCAGATTTG ATGGCTAA 1008
 (9) INFORMATION FOR SEQ ID NO:8:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 335 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Glu Ser Ser Phe Ser Phe Gly Val Ile Leu Ala Val Leu Ala Ser 15
 1
 Leu Ile Ile Ala Thr Asn Thr Leu Val Ala Val Ala Val Leu Leu 25
 20
 Ile His Lys Asn Asp Gly Val Ser Leu Cys Phe Thr Leu Asn Leu Ala 35
 40
 Val Ala Asp Thr Leu Ile Gly Val Ala Ile Ser Gly Leu Leu Thr Asp 55
 60
 Glu Leu Ser Ser Pro Ser Arg Pro Thr Glu Lys Thr Leu Cys Ser Leu 75
 80
 Arg Met Ala Phe Val Thr Ser Ser Ala Ala Ser Val Leu Thr Val 85
 90
 Met Leu Ile Thr Phe Asp Arg Tyr Leu Ala Ile Lys Glu Pro Phe Arg 100
 105
 Tyr Leu Lys Ile Met Ser Gly Phe Val Ala Gly Ala Cys Ile Ala Gly 115
 120
 Leu Trp Leu Val Ser Tyr Leu Ile Gly Phe Leu Pro Leu Gly Ile Pro 130
 135
 Met Phe Glu Glu Thr Ala Tyr Lys Gly Glu Cys Ser Phe Ala Val

- 11 -

145 150 155 160
 Phe His Pro His Phe Val Leu Thr Leu Ser Cys Val Gly Phe Phe Pro 165
 170
 Ala Met Leu Leu Phe Val Phe Phe Tyr Cys Asp Met Leu Lys Ile Ala 180
 185
 Ser Met His Ser Glu Glu Ile Arg Lys Met Glu His Ala Gly Ala Met 195
 200
 Ala Gly Gly Tyr Arg Ser Pro Arg Thr Pro Ser Asp Phe Lys Ala Leu 210
 215
 Arg Thr Val Ser Val Leu Ile Gly Ser Phe Ala Leu Ser Trp Thr Pro 225
 230
 Phe Leu Ile Thr Gly Ile Val Glu Val Ala Cys Glu Glu Cys His Leu 245
 250
 Tyr Leu Val Leu Glu Arg Tyr Leu Trp Leu Leu Gly Val Gly Asn Ser 260
 265
 Leu Leu Asn Pro Leu Ile Tyr Ala Tyr Trp Glu Lys Glu Val Arg Leu 275
 280
 Glu Leu Tyr His Met Ala Leu Gly Val Lys Lys Val Leu Thr Ser Phe 290
 295
 Leu Leu Phe Leu Ser Ala Arg Asn Cys Gly Pro Glu Arg Pro Arg Glu 305
 310
 Ser Ser Cys His Ile Val Thr Ile Ser Ser Ser Glu Phe Asp Gly 325
 330
 (10) INFORMATION FOR SEQ ID NO:9:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1413 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGAGACTA CCATGAGAGC TGACTGGGT GCCACTGGCC AAGAGCCCG CACAGACTT 60
 GATGATGAGG ACTCTTACC CCAAGTGGC TGAGACACGG TCTTCTGAT GAGCCCTGCTG 120
 CTCCTTGGGC TGACAGCCAA TGAGTTATG GCGTGGCTGG CCGGCTCCCA GAGCCGGCAT 180
 35 GAGCTGGCA CGGCTGTGCG GCTGCTCTCTG CTCAGCCTGG CCTCTCTGGA CTCTGTGTC 240

- 12 -

CTGGCAGCAG CGGCTTCCA GATCCGATG ATCCGGCATG GGGGACACTG GCCCGTGGG 300
 ACAGCTGCCT GCGCGTTCTA CTACTTCCTA TGGGGCGTGT CTTACTCTTC CGGCTCTTC 360
 CTGCTGGCCG CCTTCAGCCT CGACGGTTCG CTGCTGGCCG TGTGCCACA CTGTTACCTT 420
 GGGCACCAGC CAGTCCGCTT GCGCTCTGG GTCTGGCCG GTGCTGGGT GCTGGCCACA 480
 5 CTCTTCAGCG TGCCCTGGCT GGTCTTCCC GAGGCTGCCG TCTGTGTGTA CGACTGTGTC 540
 ATCTGCCCTGG ACTTCTGGGA CAGCGAGGAG CTGTGCTGTA GGATGCTGGA GGTCTGGGG 600
 GGTCTCTGC CTCTCTCTCT GCTGCTGTC TGCCAGTGC TCACCCAGGC CACAGCCTGT 660
 CGCACTTGC ACCGCGACA GGAGCCGCA GCCTGCCGGG GCTTGGCCCG TGTGGCCAGG 720
 ACCATTCTGT CAGCCTATGT GGTCTGAGG CTGCGCTTACC AGCTGGCCCA GCTGCTCTAC 780
 10 CTGCGCTTCC TGTGGAGCT CTACTGTGC TACCTGCTCT GGGAGGCCCT GGTCTACTCC 840
 GACTTACTGA TCCTACTCAA CAGTGGCTC AGCCCTTCC TGTGCTCTAT GGCAGTGC 900
 GACCTCCGGA CCTGCTGCG CTCCGTGTC TCGTCTTTC GCGCAGCTCT CTGGAGGAG 960
 CGGCGGGCA GCTTCAGCC CACTGAGCA CAGACCCAGC TAGATTCTGA GGTTCCTCACT 1020
 CTGCCAGAGC CGATGGCAGA GCGCCAGTCA CAGATGNTC CTGTGGCCCA GCTCAGGTG 1080
 15 AACCCACAC TCCAGCCAGC ATCGATPCC ACAGCTCAGC CACAGCTGAA CCTTACGGCC 1140
 CAGCCAGCT CGATGCCAC AGCCAGCCA CAGCTGAACC TCATGGCCCA GCCACAGTCA 1200
 GATTCTGTGG CCCAGCCACA GGCAGACACT AAGTCCAGA CCCCTGCACC TGTGCGCAGT 1260
 TCTGTGCCCA GTCCCTGTGA TGAAGCTTCC CCAACCCCAT CCTCGCATCC TACCCGAGG 1320
 GCGCTTGAG ACCAGCCAC ACTCTCTGCC TCTGAAGGAG AAGCCCCAG CAGCACCAGG 1380
 20 CCAGAGGCG CCCGCGCGC AGGCCCCAG TGA 1413

(11) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 468 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Asp Thr Thr Met Glu Ala Asp Leu Gly Ala Thr Gly His Arg Pro
 1 5 10 15

- 13 -

Arg Thr Glu Leu Asp Asp Glu Asp Ser Tyr Pro Gln Gly Gly Tyr Asp
 20 25 30
 Thr Val Phe Leu Val Ala Leu Leu Leu Gly Leu Pro Ala Asn Gly
 35 40 45
 5 Leu Met Ala Tyr Leu Ala Gly Ser Gln Ala Arg His Gly Ala Gly Thr
 50 55 60
 Arg Leu Ala Leu Leu Ser Leu Ala Leu Ser Asp Phe Leu Phe
 65 70 75 80
 Leu Ala Ala Ala Phe Gln Ile Leu Glu Ile Arg His Gly Gly His
 85 90 95
 10 Trp Pro Leu Gly Thr Ala Ala Cys Arg Phe Tyr Tyr Phe Leu Trp Gly
 100 105 110
 Val Ser Tyr Ser Ser Gly Leu Phe Leu Leu Ala Ala Leu Ser Leu Asp
 115 120 125
 15 Arg Cys Leu Leu Ala Leu Cys Pro His Trp Tyr Pro Gly His Arg Pro
 130 135 140
 Val Arg Leu Pro Leu Trp Val Cys Ala Gly Val Trp Val Leu Ala Thr
 145 150 155 160
 Leu Phe Ser Val Pro Trp Leu Val Phe Pro Glu Ala Ala Val Trp Trp
 165 170 175
 Tyr Asp Leu Val Ile Cys Leu Asp Phe Trp Asp Ser Glu Glu Leu Ser
 180 185 190
 Leu Arg Met Leu Glu Val Leu Gly Gly Phe Leu Pro Phe Leu Leu Leu
 195 200 205
 25 Leu Val Cys His Val Leu Thr Gln Ala Thr Arg Thr Cys His Arg Gln
 210 215 220
 Gln Gln Pro Ala Ala Cys Arg Gly Phe Ala Arg Val Ala Arg Thr Ile
 225 230 235 240
 30 Leu Ser Ala Tyr Val Val Leu Arg Leu Pro Tyr Gln Leu Ala Gln Leu
 245 250 255
 Leu Tyr Leu Ala Phe Leu Trp Asp Val Tyr Ser Gly Tyr Leu Leu Trp
 260 265 270
 Glu Ala Leu Val Tyr Ser Asp Tyr Leu Ile Leu Leu Asn Ser Cys Leu
 275 280 285
 35 Ser Pro Phe Leu Cys Leu Met Ala Ser Ala Asp Leu Arg Thr Leu Leu
 290 295 300
 Arg Ser Val Leu Ser Ser Phe Ala Ala Leu Cys Glu Arg Pro

- 14 -

305 310 315 320
 Gly Ser Phe Thr Pro Thr Glu Pro Gln Thr Gln Leu Asp Ser Glu Gly
 325 330 335
 Pro Thr Leu Pro Glu Pro Met Ala Glu Ala Gln Ser Gln Met Asp Pro
 340 345 350
 Val Ala Gln Pro Gln Val Asn Pro Thr Leu Gln Pro Arg Ser Asp Pro
 355 360 365
 Thr Ala Gln Pro Gln Leu Asn Pro Thr Ala Gln Pro Gln Ser Asp Pro
 370 375 380
 Thr Ala Gln Pro Gln Leu Asn Leu Met Ala Gln Pro Gln Ser Asp Ser
 385 390 395 400
 Val Ala Gln Pro Gln Ala Asp Thr Asn Val Gln Thr Pro Ala Pro Ala
 405 410 415
 Ala Ser Ser Val Pro Ser Pro Cys Asp Glu Ala Ser Pro Thr Pro Ser
 420 425 430
 Ser His Pro Thr Pro Gly Ala Leu Glu Asp Pro Ala Thr Pro Pro Ala
 435 440 445
 Ser Glu Gly Glu Ser Pro Ser Ser Thr Pro Pro Glu Ala Ala Pro Gly
 450 455 460
 Ala Gly Pro Thr
 465

(12) INFORMATION FOR SEQ ID NO:11:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1248 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:
 30 AATGACGAGA TGAATAAAT TCAGATGCT TCCGTGATC ACCACAGAA ACTAAGAT 60
 CCATTCAGA AACACCTGA CAGACCAAG GAGTATCTGG CCTTCCTCTG CGAGCTCGG 120
 CGCAGCACT TCTTCTCTCC CGTGTCTGG GTGTATGTGC CAATTTTGT GGTGGGGGTC 180
 ATTGGCAATG TCCGTGTGG CTTGTGATC CTGACAGAC AGGCTATGAA GACGCCACAC 240
 AACTACTACC TCTTCAGCT GGGGTCTCT GACCTCTCG TCCGTCTCT TGAATGCC 300

- 15 -

CTGAGGCTCT ATGAGATGTG GCGCAACTAC CCTTTCTTGT TCGGACCGGT GGGCTGTAC 360
 TTCAGACAG CCCTCTTTGA GACCGTGTGC TTGCGCTTCA TCTTCAGAT CACACCGTC 420
 AGCGTGAAC GCTACGTGAC CATCTTACAC CGGTTCGGCG CCAACTGCA GAGCACCGG 480
 CGCGGGGCC TCAAGATCTT CGGATCTGTC TGGGGCTTCT CGGTCTCTT CTCCCTGCC 540
 5 AACACAGCA TCCATGGCAT CAGTTTCAC TACTTCCCA ATGGTCTCT GGTCCAGGT 600
 TCGGCCACCT GTACGTCAT CAGGCCATG TGAATCTACA ATTTCATAT CAGGTTCACC 660
 TCTTCTCTAT TCTACTCTCT CCCCATGAT GTCATCATG TCCCTTACTA CTTCATGGCA 720
 CTCAGACTAA AGAAAGACA ATCTCTTGA GCAATGAAAG GAATGCATA TATCAAGA 780
 CCTTCAGAAA AATCAGTCA CAGATGCTG TTGTCTTGG TCTTATGTT TGTATCTGT 840
 10 TGGGCCCCGT TCCAGATTGA CGACTCTTC TTCACTTTG TGGAGAGTG GAGTGAATCC 900
 CTGCTGTCTG TGTTCACCT GATCCATGAG GTGTCAAGTG TCTTCTTCTA CTTGAGCTCA 960
 GCTGTCAAC CCATATCTA TAACTACTG TCTGCGGCT TCCAGGACG ATTCAGAAAT 1020
 GTATCTCTT CTTTCACAA ACAATGACG TCCAGAGTG ACCCAAGTT GCCACTGAC 1080
 CAGCGAACA TCTTCTTAC ABAATGCCAC TTGTGAGC TGACCGAAGA TATAGTCCC 1140
 15 CAATTCGAT GTCAATCAT CAGCAGAAC TCTCACTCC CACAGCCCT CTCTAGTAA 1200
 CAGATGTCA GAACAATA TCAAGCTTC CACTTACA AACTGAA 1248

(13) INFORMATION FOR SEQ ID NO:12:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 415 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:12:
 25 Met Ser Gly Met Glu Lys Leu Gln Asn Ala Ser Trp Ile Tyr Gln Gln
 1 5 10 15
 Lys Leu Glu Asp Pro Phe Gln Lys His Leu Asn Ser Thr Glu Glu Tyr
 20 25 30
 Leu Ala Phe Leu Cys Gly Pro Arg Arg Ser His Phe Phe Leu Pro Val
 35 40 45
 Ser Val Val Tyr Val Pro Ile Phe Val Val Gly Val Ile Gly Asn Val

- 16 -

50 55 60
 Leu Val Cys Leu Val Ile Leu Gln His Gln Ala Met Lys Thr Pro Thr 80
 65 70 75
 Asn Tyr Tyr Leu Phe Ser Leu Ala Val Ser Asp Leu Leu Val Leu Leu 95
 85 90
 Leu Gly Met Pro Leu Glu Val Tyr Glu Met Trp Arg Asn Tyr Pro Phe 110
 100 105 110
 Leu Phe Gly Pro Val Gly Cys Tyr Phe Lys Thr Ala Leu Phe Glu Thr 125
 115 120 125
 Val Cys Phe Ala Ser Ile Leu Ser Ile Thr Thr Val Ser Val Glu Arg 140
 130 135 140
 Tyr Val Ala Ile Leu His Pro Phe Arg Ala Lys Leu Gln Ser Thr Arg 155
 145 150 155 160
 Arg Arg Ala Leu Arg Ile Leu Gly Ile Val Trp Gly Phe Ser Val Leu 175
 165 170 175
 Phe Ser Leu Pro Asn Thr Ser Ile His Gly Ile Lys Phe His Tyr Phe 190
 180 185 190
 Pro Asn Gly Ser Leu Val Pro Gly Ser Ala Thr Cys Thr Val Ile Lys 205
 195 200 205
 Pro Met Trp Ile Tyr Asn Phe Ile Ile Gln Val Thr Ser Phe Leu Phe 220
 210 215 220
 Tyr Leu Leu Pro Met Thr Val Ile Ser Val Leu Tyr Tyr Leu Met Ala 240
 225 230 235 240
 Leu Arg Leu Lys Lys Asp Lys Ser Leu Glu Ala Asp Gly Asn Ala 255
 245 250 255
 Asn Ile Gln Arg Pro Cys Arg Lys Ser Val Asn Lys Met Leu Phe Val 270
 260 265 270
 Leu Val Leu Val Phe Ala Ile Cys Trp Ala Pro Phe His Ile Asp Arg 285
 275 280 285
 Leu Phe Phe Ser Phe Val Glu Glu Trp Ser Glu Ser Leu Ala Ala Val 300
 290 295 300
 Phe Asn Leu Val His Val Val Ser Gly Val Phe Phe Tyr Leu Ser 320
 305 310 315 320
 Ala Val Asn Pro Ile Ile Tyr Asn Leu Leu Ser Arg Arg Phe Gln Ala 335
 325 330 335
 Ala Phe Gln Asn Val Ile Ser Ser Phe His Lys Gln Trp His Ser Gln 350
 340 345 350

- 17 -

His Asp Pro Gln Leu Pro Pro Ala Gln Arg Asn Ile Phe Leu Thr Glu 355
 360 365
 Cys His Phe Val Glu Leu Thr Glu Asp Ile Gly Pro Gln Phe Pro Cys 380
 370 375 380
 5 Gln Ser Ser Met His Asn Ser His Leu Pro Thr Ala Leu Ser Ser Glu 395
 385 390 395 400
 Gln Met Ser Arg Thr Asn Tyr Gln Ser Phe His Phe Asn Lys Thr 415
 405 410 415
 (14) INFORMATION FOR SEQ ID NO:13:
 10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1173 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 15 (ii) MOLECULE TYPE: DNA (genomic)
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
 ATGCCAGATA CTAATAGCAC ATCAATTTA TCACTAAGCA CTCGTGTTAC TTAGCATTT 60
 TTATGTCTT TAGTAGCTTT TGCTATAATG CTAGGAATG CTTGGTCTAT TTAGCTTTT 120
 GTGGTGGACA AAACCTTAG ACATCGAAGT AGTTATTTTT TTCTTAACCT GGCCATCTT 180
 20 GACTTCTTTG TGGGTGTGAT CTCCTTCCT TTGTACATCC CTCACACGCT GTTCGAATGG 240
 GATTTTGGAA AGGAATCTG TGTATTTTGG CTCACTACTG ACTATCTGTT ATGTACAGCA 300
 TCTGTATATA ACATGTCTCT CATCAGCTAT GATCGATACC TGTCACTCTC AATGCTGTG 360
 TCTTATAGAA CTCACATAC TGGGTCTTG AAGATTGTTA CTCTGATGCT GGCCGTTTGG 420
 GTGCTGGCCT TCTTAGTGAA TGGGCCAATG ATTCTAGTTT CAGAGTCTTG GAAGGATGAA 540
 25 GGTAGTGAAT GTCAACCTGG ATTTTTTTCG GAATGGTACA TCCTTGCCAT CACATCATTC 540
 TTGGAAATCG TGAATCCAGT CATCTAGTC GCTTATTCA ACATGAATAT TTATTGGAGC 600
 CTGTGGAGCG GTGATCATCT CAGTAGGTGC CAAAGCCATC CTGGACTGAC TGCTGTCTCT 660
 TCCAACATCT GTGGACACTC ATTCAAGAGT AGACTATCTT CAAGGAGATC TCTTTCTGCA 720
 TCACACAGAG TTCTTGCATC CTTTCATTTCA GAGAGACAGA GGAGAAGAG TAGTCTCATG 780
 30 TTTTCTCTCA GAACCAAGAT GAATAGCAAT ACAATTGCTT CCAAAATGGG TTCCTTCTCC 840
 CAATCAGATT CTGTAGCTCT TCACCAAGG GAACATGTTG AACTGCTTAG AGCCAGAGAA 900

- 18 -

TTAGCCAAAGT CACTGGCCAT TCTCTTAGGG GTTTTGGCTG TTGGCTGGGC TCCCAATTC 960
 CTGTGCACA TTGTCTCTTC ATTATATCC TCAGCAACAG GTCCCTAAATC AGTTGGTAT 1020
 AGAATTGCAT TTGGGCTCA GTGGTGCAT TCCTTGTCAT ATCCTCTTT GTATCCATGG 1080
 TGTCAACAGC GCTTCAAAA GCGTTCTTG AAAATATTT GTATAAAAA GCACCTCTA 1140
 CCATCACAC ACAGTGGTC AGATCTTCT TAA 1173

(15) INFORMATION FOR SEQ ID NO:14:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 390 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Pro Asp Thr Asn Ser Thr Ile Asn Leu Ser Leu Ser Thr Arg Val 1
 5
 Thr Leu Ala Phe Phe Met Ser Leu Val Ala Phe Ala Ile Met Leu Gly 20
 25
 Asn Ala Leu Val Ile Leu Ala Phe Val Val Asp Lys Asn Leu Arg His 35
 40
 Arg Ser Ser Tyr Phe Phe Leu Asn Leu Ala Ile Ser Asp Phe Phe Val 50
 55
 Gly Val Ile Ser Ile Pro Leu Tyr Ile Pro His Thr Leu Phe Glu Trp 65
 70
 Asp Phe Gly Lys Glu Ile Cys Val Phe Trp Leu Thr Thr Asp Tyr Leu 85
 90
 Leu Cys Thr Ala Ser Val Tyr Asn Ile Val Leu Ile Ser Tyr Asp Arg 100
 105
 Tyr Leu Ser Val Ser Asn Ala Val Ser Tyr Arg Thr Gln His Thr Gly 115
 120
 Val Leu Lys Ile Val Thr Leu Met Val Ala Val Trp Val Leu Ala Phe 130
 135
 Leu Val Asn Gly Pro Met Ile Leu Val Ser Glu Ser Trp Lys Asp Glu 145
 150
 Gly Ser Glu Cys Glu Pro Gly Phe Phe Ser Glu Trp Tyr Ile Leu Ala 165
 170

- 19 -

Ile Thr Ser Phe Leu Glu Phe Val Ile Pro Val Ile Leu Val Ala Tyr 180
 185
 Phe Asn Met Asn Ile Tyr Trp Ser Leu Trp Lys Arg Asp His Leu Ser 195
 200
 Arg Cys Gln Ser His Pro Gly Leu Thr Ala Val Ser Ser Asn Ile Cys 210
 215
 Gly His Ser Phe Arg Gly Arg Leu Ser Ser Arg Arg Ser Leu Ser Ala 225
 230
 Ser Thr Glu Val Pro Ala Ser Phe His Ser Glu Arg Gln Arg Lys 245
 250
 Ser Ser Leu Met Phe Ser Ser Arg Thr Lys Met Asn Ser Asn Thr Ile 260
 265
 Ala Ser Lys Met Gly Ser Phe Ser Gln Ser Asp Ser Val Ala Leu His 275
 280
 Gln Arg Glu His Val Glu Leu Leu Arg Ala Arg Arg Leu Ala Lys Ser 290
 295
 Leu Ala Ile Leu Leu Gly Val Phe Ala Val Cys Trp Ala Pro Tyr Ser 305
 310
 Leu Phe Thr Ile Val Leu Ser Phe Tyr Ser Ser Ala Thr Gly Pro Lys 325
 330
 Ser Val Trp Tyr Arg Ile Ala Phe Trp Leu Gln Trp Phe Asn Ser Phe 340
 345
 Val Asn Pro Leu Leu Tyr Tyr Pro Leu Cys His Lys Arg Phe Gln Lys Ala 355
 360
 Phe Leu Lys Ile Phe Cys Ile Lys Lys Gln Pro Leu Pro Ser Gln His 370
 375
 Ser Arg Ser Val Ser Ser 385
 390

(16) INFORMATION FOR SEQ ID NO:15:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGAAAGCTTA ACGATCCCA GGAGCAACAT

30

(17) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CTGGATCCT ACGAGACAT TTTTCACACA G
31

(18) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1128 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

ATGCGAAGC CGAGCAGCC GGGTGGCAGC GCGGCGCGC AGCGCGCCG CCTGGGCCTC 60
 AAGCTGGCCA CGCTCAGCT GCTGCTGTGC GTGAGCCTAG CCGGCACAGT GCTGTTGCGC 120
 CTGCTGATCG TCGGGGAGCG CAGCTGACAC CGCGCCCGCT ACTACCTGCT GCTCGACCTG 180
 TGCTTGCCG ACGGGCTGCG CGGCTCGCC TGCTCCCGG CCGTCATGCT GCGGCGCGCG 240
 CGTGGCGCG CGCGCGCGG GCGCGCGCG GCGCGCTGG GCTGCAAGCT GCTCGCCTTC 300
 CTGGCGCGC TCCTCTGCTT CCAGCGCGC TTCTGCTGC TGGGCGTGG CGTCAACCGC 360
 TACCTGGCCA TCGCGCACCA CCGCTTCTAT GCAGAGCCG TGGCGGCGTG GCGGTGCGCC 420
 GCATGCTGG TGTGGCGCGC CTGGGCGCTG GCCTGCGCG CCGCTTCTCC GCCAGTGTG 480
 GACGGCGGTG GCGACGACCA GGAGCGCGCG TGGCGCCCTG AGCAGCGGCC CGACGCGGCC 540
 CCGCGCGCG TGGGCTTCTT GCTGCTGCTG CCGGTGGTGG TGGGCGCCAC GCACCTCGTC 600
 TACCTCGGCC TGCTCTTCTT CATCCACGAC CGCGCGCAGA TGGCGCGCG GCGCTGGTGG 660

CCGCGCTCA GCCAGACTG GACCTTCCAC GSCCCGGGCG CCACCGGCCA GCGCGCGGCC 720
 AACTGGACGG CCGGCTTCGG CCGCGGCGCC ACGCGCGCG CGCTTGTGGG CATCGGCCCC 780
 GCAGGCGCG GCGCGGCGCG GCGCGGCGCT CTCGTGTGCG AAGAAATCAA GACGAGAGAAG 840
 AGGCTGTGCA AGATGTTCTA CCGCGTCACG CTGCTCTTCC TGCTCTCTCG GGGGCGCTAC 900
 5 GTGCTGGCCA GCTACCTGCG GGTCTGTGTT CCGCGCGCG CCGTCCGCCA GGCCTACCTG 960
 ACGGCTCCG TGTGGCTGAC CTTGCGGCG GCGGCGCATCA ACCCGCTCGT GTGCTTCTTC 1020
 TTCAACAGGG AGCTGAGGA CTGCTTCAGG GCCAGTTCCT CTTGCTGGCA GAGCCCCCGG 1080
 ACCACCCAGG CGACCCATCC CTGCGACCTG AAAGGCATTG GTTTATGA

(19) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 375 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Ala Asn Ala Ser Glu Pro Gly Gly Ser Gly Gly Gly Glu Ala Ala 15
 1 5 10
 Ala Leu Gly Leu Lys Leu Ala Thr Leu Ser Leu Leu Cys Val Ser 30
 20 25 30
 Leu Ala Gly Asn Val Leu Phe Ala Leu Leu Ile Val Arg Glu Arg Ser 45
 35 40 45
 Leu His Arg Ala Pro Tyr Tyr Leu Leu Leu Asp Leu Cys Leu Ala Asp 60
 50 55 60
 Gly Leu Arg Ala Leu Ala Cys Leu Pro Ala Val Met Leu Ala Ala Arg 80
 65 70 75
 Arg Ala Ala Ala Ala Gly Ala Pro Pro Gly Ala Leu Gly Cys Lys 95
 85 90 95
 Leu Leu Ala Phe Leu Ala Leu Phe Cys Phe His Ala Ala Phe Leu 110
 100 105 110
 Leu Leu Gly Val Gly Val Thr Arg Tyr Leu Ala Ile Ala His His Arg 125
 115 120 125
 Phe Tyr Ala Glu Arg Leu Ala Gly Trp Pro Cys Ala Ala Met Leu Val 140
 130 135 140

- 22 -

Cys Ala Ala Trp Ala Leu Ala Leu Ala Ala Phe Pro Pro Val Leu
 145 150 155 160
 Asp Gly Gly Gly Asp Asp Glu Asp Ala Pro Cys Ala Leu Glu Gln Arg
 165 170 175
 Pro Asp Gly Ala Pro Gly Ala Leu Gly Phe Leu Leu Leu Ala Val
 180 185 190
 Val Val Gly Ala Thr His Leu Val Tyr Leu Arg Leu Leu Phe Ile
 195 200 205
 His Asp Arg Arg Lys Met Arg Pro Ala Arg Leu Val Pro Ala Val Ser
 210 215 220
 His Asp Trp Thr Phe His Gly Pro Gly Ala Thr Gly Gln Ala Ala
 225 230 235 240
 Asn Trp Thr Ala Gly Phe Gly Arg Gly Pro Thr Pro Ala Leu Val
 245 250 255
 Gly Ile Arg Pro Ala Gly Pro Gly Arg Gly Ala Arg Arg Leu Leu Val
 260 265 270
 Leu Glu Glu Phe Lys Thr Glu Lys Arg Leu Cys Lys Met Phe Tyr Ala
 275 280 285
 Val Thr Leu Leu Phe Leu Leu Leu Trp Gly Pro Tyr Val Val Ala Ser
 290 295 300
 Tyr Leu Arg Val Leu Val Arg Pro Gly Ala Val Pro Gln Ala Tyr Leu
 305 310 315 320
 Thr Ala Ser Val Trp Leu Thr Phe Ala Gln Ala Gly Ile Asn Pro Val
 325 330 335
 Val Cys Phe Leu Phe Asn Arg Glu Leu Arg Asp Cys Phe Arg Ala Gln
 340 345 350
 Phe Pro Cys Cys Gln Ser Pro Arg Thr Thr Gln Ala Thr His Pro Cys
 355 360 365
 Asp Leu Lys Gly Ile Gly Leu
 370 375

(20) INFORMATION FOR SEQ ID NO:19:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1002 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)

- 23 -

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:19:
 ATGACACCA CAGTATGCA AGGCTGAC AGATGAGC GGTGCCCG AGACACTGG 60
 ATGATACAG TGGTATGCC AGCCCTGAC AGAGTGGT TCTGACCG CATCTGCTG 120
 AATACCTGG CTCTGAGGT GTTGTGAC ATCCGAGCT CCGCACCCT CATCATCTAC 180
 CTCAAAACA CTTGGTGGC GCACCTGATA ATGACACTGA TGCCTTCCTT CAAATCTTC 240
 TCTGACTGAC ACCTGGCACC CTGGCAGCTC AGAGCTTTTG TGTGTCGTTT TTCTTGGGTG 300
 ATATTTATG AGACCAATGA TGGGGCATC GTGCTGTATG GGTCTATGAC CTTTGACAGA 360
 TTCTTCAGG TGCATGACC TTGAGAAAT ATTTTCTTA AAAACCTGT TTTTGCAAAA 420
 ACGGTCTGAA TCTTCACTG GTTCTTTTG TCTTCATCT CCGTCGAAA TACGATCTTG 480
 AGCAACAAG AAGCAACCC ATGCTCTGAG AAAAGTGTG CTTCCTTAA GGGGCTCTG 540
 GGGCTGAAT GGCATCAAA GGTAAATAC ATAGCCAGT TTAATTTCTG GACTGTTTT 600
 ATCTAATGC TTGTGTTTTA TGTGTTAT GTAAAAAG TATATGATTC TTAAGAAAAG 660
 TCCAAAGTA AAGACAGAAA AAACACAAA AAGCTGAGAG GCAAGTAT TTGTGCTG 720
 GCTGTCTCT TTGTGTGTT TGTCTCATTT CATTTGCCA GAGTTCATA TACTGCAGT 780
 CAAACCAACA ATTAAGCTGA CTGTAGCTG CAATATCAC TGTATTATG TAAAGAAACA 840
 ACTCTCTTT TGGCAGCAGC TAACTTTGT ATGATCCCT TAAATACAT ATCTTATGT 900
 AAAAAATCA CAGAAAGCT ACCATGTATG CAAGGAGAA AGACCAAGC ATCAAGCAA 960
 GAATATCAT GCACTGAGC AGACAACTA ACCTTAGCT GA 1002

(21) INFORMATION FOR SEQ ID NO:20:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 333 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (11) MOLECULE TYPE: protein

Met Asn Thr Thr Val Met Gln Gly Phe Asn Arg Ser Glu Arg Cys Pro
 1 5 10 15
 Arg Asp Thr Arg Ile Val Gln Leu Val Phe Pro Ala Leu Tyr Thr Val
 20 25 30

-24-

Val Phe Leu Thr Gly Ile Leu Leu Asn Thr Leu Ala Leu Trp Val Phe
35 40 45

Val His Ile Pro Ser Ser Thr Phe Ile Ile Tyr Leu Lys Asn Thr
50 55 60

Leu Val Ala Asp Leu Ile Met Thr Leu Met Leu Pro Phe Lys Ile Leu
65 70 75 80

Ser Asp Ser His Leu Ala Pro Trp Gln Leu Arg Ala Phe Val Cys Arg
85 90 95

Phe Ser Ser Val Ile Phe Tyr Glu Thr Met Tyr Val Gly Ile Val Leu
100 105 110 115

Leu Gly Leu Ile Ala Phe Asp Arg Phe Leu Lys Ile Ile Arg Pro Leu
115 120 125

Arg Asn Ile Phe Leu Lys Lys Pro Val Phe Ala Lys Thr Val Ser Ile
130 135 140

Phe Ile Trp Phe Phe Leu Phe Phe Ile Ser Leu Pro Asn Thr Ile Leu
145 150 155 160

Ser Asn Lys Glu Ala Thr Pro Ser Val Lys Lys Cys Ala Ser Leu
165 170 175

Lys Gly Pro Leu Gly Leu Lys Trp His Gln Met Val Asn Asn Ile Cys
180 185 190

Gln Phe Ile Phe Trp Thr Val Phe Ile Leu Met Leu Val Phe Tyr Val
195 200 205

Val Ile Ala Lys Lys Val Tyr Asp Ser Tyr Arg Lys Ser Lys Ser Lys
210 215 220

Asp Arg Lys Asn Asn Lys Lys Leu Glu Lys Val Phe Val Val Val
225 230 235 240

Ala Val Phe Phe Val Cys Phe Ala Pro Phe His Phe Ala Arg Val Pro
245 250 255

Tyr Thr His Ser Gln Thr Asn Asn Lys Thr Asp Cys Arg Leu Gln Asn
260 265 270

Gln Leu Phe Ile Ala Lys Glu Thr Thr Leu Phe Leu Ala Ala Thr Asn
275 280 285

Ile Cys Met Asp Pro Leu Ile Tyr Ile Phe Leu Cys Lys Lys Phe Thr
290 295 300

Glu Lys Leu Pro Cys Met Gln Gly Arg Lys Thr Thr Ala Ser Ser Gln
305 310 315 320

Glu Asn His Ser Ser Gln Thr Asp Asn Ile Thr Leu Gly

-25-

325 330

(22) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1122 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:

10 ATGGCCACACA CTCACGGAGAGA GCTGAGGAG GTGAGGGGG CTCTGTCCCC ACCGTCCGCA
TCAGCTTATG TGAAGCTGCT ACTGCTGGGA CTGATTATGT GGCTGAGCCT GGCAGGTAAAC 120
GCCATCTTGT CCTGCTGGT GCTCAAGGAG CGTGCCTGCG ACAAGGCTCC TTACTACTTC 180
CTGCTGAGCC TGTGCTGGC CGATGGCATA CGCTCTGGCG TCTGCTTCCC CTTTGTGCTG 240
GCTTCTGTC GCACGGGCTC TTCATGGACC TTCAGTGCAC TCAGCTGCAC GATTGTGCCC 300
TTTATGGCCG TGCTCTTTTG CTTCATGGCG GCCTTCATGC TGTTCTGCAT CAGCGTCACC 360
CGCTACATGG CCATGCGCCA CCACGGCTTC TACGCCAAGC GCATGACACT CTGACATGCG 420
GGCGCTGTCA TCTGCATGGC CTGGACCCCTG TCTGTGGCCA TGGCCTTCCC ACCTGTCTTT 480
GACGTGGGCA CCTACAAATT TATTGGGGAG GAGGACCAGT GCATCTTTGA GCATCGCTAC 540
TTCAAGGCCA ATGACACGCT GGGCTTTCATG CTTATGTTGG CTGTGCTCAT GGCAGCTACC 600
20 CATGCTGTCT ACGGCAAGCT GCTCTCTTC GAGTATCGTC ACCGCAAGAT GAAGCCAGTG 660
CAGATGCTGC CAGCCATCAG CCAGAACTGG ACATTCCATG GTCCCGGGGC CACCGCCAG 720
GCTGCTGCCA ACTGGATCGC CGGCTTTGGC CGTGGGCCCA TGCACCAAC CTTGCTGGGT 840
ATCCGGCAGA ATGGGCATGC AGCCAGCCGG CGGCTACTGG GCATGGACGA GGTCAAGGGT 900
GAAAAGCAGC TGGGCCCGCAT GTTCTACGCG ATCACACTGC TCTTTCGTCT CTTCTGCTCA 960
25 CCCTACATCG TGGCTGCTTA CTGGCGAGTG TTGTGAAG CCTGTGCTGT GCGCCACCGC 1020
TACCTGSCCA CTGCTGTTTG GATGAGCTTC GCCCAGGCTG CCGTCAACCC AATTGTCTGC 1080
TTCTGTGCTCA ACAAGGACCT CAAGAAGTGC CTGACCACCTC ACGCCCCCTG CTGGGGCACA 1122
GGAGTGCCCG CGGTCTCCAG AGAACCCCTAC TGTGTATGT GA

(23) INFORMATION FOR SEQ ID NO:22:

- 26 -

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 373 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Met Ala Asn Thr Thr Gly Glu Pro Glu Glu Val Ser Gly Ala Leu Ser
1 5 10 15
Pro Pro Ser Ala Ser Ala Tyr Val Lys Leu Val Leu Gly Leu Ile
20 25 30
Met Cys Val Ser Leu Ala Gly Asn Ala Ile Leu Ser Leu Leu Val Leu
35 40 45
Lys Glu Arg Ala Leu His Lys Ala Pro Tyr Tyr Phe Leu Leu Asp Leu
50 55 60
Cys Leu Ala Asp Gly Ile Arg Ser Ala Val Cys Phe Pro Phe Val Leu
65 70 75 80
Ala Ser Val Arg His Gly Ser Ser Trp Thr Phe Ser Ala Leu Ser Cys
85 90 95
Lys Ile Val Ala Phe Met Ala Val Leu Phe Cys Phe His Ala Ala Phe
100 105 110
Met Leu Phe Cys Ile Ser Val Thr Arg Tyr Met Ala Ile Ala His His
115 120 125
Arg Phe Tyr Ala Lys Arg Met Thr Leu Trp Thr Cys Ala Ala Val Ile
130 135 140
Cys Met Ala Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Phe
145 150 155 160
Asp Val Gly Thr Tyr Lys Phe Ile Arg Glu Glu Asp Glu Cys Ile Phe
165 170 175
Glu His Arg Tyr Phe Lys Ala Asn Asp Thr Leu Gly Phe Met Leu Met
180 185 190
Leu Ala Val Leu Met Ala Ala Thr His Ala Val Tyr Gly Lys Leu Leu
195 200 205
Leu Phe Glu Tyr Arg His Arg Lys Met Lys Pro Val Gln Met Val Pro
210 215 220
Ala Ile Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Thr Gly Gln
225 230 235

- 27 -

Ala Ala Ala Asn Trp Ile Ala Gly Phe Gly Arg Gly Pro Met Pro Pro
245 250 255

Thr Leu Leu Gly Ile Arg Gln Asn Gly His Ala Ala Ser Arg Arg Leu
260 265 270

Leu Gly Met Asp Glu Val Lys Gly Glu Lys Gln Leu Gly Arg Met Phe
275 280 285

Tyr Ala Ile Thr Leu Leu Phe Leu Leu Leu Trp Ser Pro Tyr Ile Val
290 295 300

Ala Cys Tyr Trp Arg Val Phe Val Lys Ala Cys Ala Val Pro His Arg
305 310 315 320

Tyr Leu Ala Thr Ala Val Trp Met Ser Phe Ala Gln Ala Ala Val Asn
325 330 335

Pro Ile Val Cys Phe Leu Leu Asn Lys Asp Leu Lys Lys Cys Leu Thr
340 345 350

Thr His Ala Pro Cys Trp Gly Thr Gly Gly Ala Pro Ala Pro Arg Glu
355 360 365

Pro Tyr Cys Val Met
370

(24) INFORMATION FOR SEQ ID NO:23:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1053 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ATGGCTTTGG AACAAACCA GTCAACAGAT TATTATTATG AGGAAATGA AATTGATGCG 60
ACTTATGACT AAGTCAATA TGATATGATC TGATATCAAG AAGATGTGCG AAGATTGCA 120
AAAGTTTCC TCCCTGATTT CCTCACAATA GCTTGTGTCA TTGACTTTC AGGCAATTC 180
AATGATATG CAATTATGC CTATTACAG AACACAGAA CCAAAACAG TGTGTACATC 240
CTGAATTTGG CTGTACAGA TTACTCTCT CTATTCATC TCCCTTTTGG GCGTGTATAT 300
GCGATTCATG GGTGGGTTT AGGAAATAA ATGTGCATA TACTTCAGC CTGTACACA 360
CTAACTTG TCTGTGAT GCGATTTCG GCTTGATCA GCATAGACG ATATGTGCA 420
GTACTATATG TCCCAACCA ATCAGAGTG GAAACCAT GCTGATCAT CTGTTCCTGT 480

- 28 -

GCTGTGATGG CTGCCATCTT GCTGAGCATG CCCAGCTGG TTTTATTATC AGTAATGAC 540
 AATGCTAGGT GCATTCCCAT TTTCGCCCGC TACCTAGGAA CATCAATGAA AGCATTGATT 600
 CAAATGCTAG AGACTGTGCAT TGGATTGTGA GTACCCCTTC TTATTATGGG GGTGTGCTAC 660
 TTTATCAGG CAGGAGCACT CATGAGATG CCAACATTA AATATCTCG ACCCTTAAAA 720
 5 GTTCTGCTCA CAGTGGTTAT AGTTTTCATT GTCACGTCAAC TGCCTTATTA CATTGTCAAG 780
 TTCTGCGGAG CCATAGACAT CATCTACTCC CTGATCACC GCTGCAACAT GAGCAACGC 840
 ATGGACATCG CCATCCCACT CACAGAAAGC ATTGCATCTT TTCACAGCTG CCTACACCCA 900
 ATCCTTTATG TTTTATGGG AGCATCTTTC AAAAACTACG TTATGAAAGT GGCCAAAGAA 960
 TATGGTCTCT GGAGAAGACA GAGACAAAGT GTGGAGGAGT TTCTTTTGA TTCTGAGGCT 1020
 10 CCTACAGAGC CAAACAGTAC TTTTACCAT TAA 1053

(25) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 350 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

Met Ala Leu Glu Gln Asn Gln Ser Thr Asp Tyr Tyr Tyr Glu Glu Asn 15
 1 5 10
 Glu Met Asn Gly Thr Tyr Asp Tyr Ser Gln Tyr Glu Leu Ile Cys Ile 30
 20 25
 Lys Glu Asp Val Arg Glu Phe Ala Lys Val Phe Leu Pro Val Phe Leu 45
 35 40
 Thr Ile Ala Phe Val Ile Gly Leu Ala Gly Asn Ser Met Val Val Ala 60
 50 55
 Ile Tyr Ala Tyr Tyr Lys Lys Gln Arg Thr Lys Thr Asp Val Tyr Ile 80
 65 70 75
 Leu Asn Leu Ala Val Ala Asp Leu Leu Leu Phe Thr Leu Pro Phe 95
 85 90
 Trp Ala Val Asn Ala Val His Gly Trp Val Leu Gly Lys Ile Met Cys 110
 100 105
 Lys Ile Thr Ser Ala Leu Tyr Thr Leu Asn Phe Val Ser Met Gln

- 29 -

115 120 125
 Phe Leu Ala Cys Ile Ser Ile Asp Arg Tyr Val Ala Val Thr Asn Val 140
 130 135
 Pro Ser Gln Ser Gly Val Gly Lys Pro Cys Trp Ile Ile Cys Phe Cys 160
 145 150 155
 Val Trp Met Ala Ala Ile Leu Leu Ser Ile Pro Gln Leu Val Phe Tyr 175
 165 170
 Thr Val Asn Asp Asn Ala Arg Cys Ile Pro Ile Phe Pro Arg Tyr Leu 190
 180 185
 Gly Thr Ser Met Lys Ala Leu Ile Gln Met Leu Glu Ile Cys Ile Gly 205
 195 200
 Phe Val Val Pro Phe Leu Ile Met Gly Val Cys Tyr Phe Ile Thr Ala 220
 210 215
 Arg Thr Leu Met Lys Met Pro Asn Ile Lys Ile Ser Arg Pro Leu Lys 240
 225 230 235
 Val Leu Leu Thr Val Val Ile Val Phe Ile Val Thr Gln Leu Pro Tyr 255
 245 250
 Asn Ile Val Lys Phe Cys Arg Ala Ile Asp Ile Ile Tyr Ser Leu Ile 270
 260 265
 Thr Ser Cys Asn Met Ser Lys Arg Met Asp Ile Ala Ile Gln Val Thr 285
 275 280
 Glu Ser Ile Ala Leu Phe His Ser Cys Leu Asn Pro Ile Leu Tyr Val 300
 290 295
 Phe Met Gly Ala Ser Phe Lys Asn Tyr Val Met Lys Val Ala Lys Lys 320
 305 310 315
 Tyr Gly Ser Trp Arg Arg Gln Arg Gln Ser Val Glu Glu Phe Pro Phe 335
 325 330
 Asp Ser Glu Gly Pro Thr Glu Pro Thr Ser Thr Phe Ser Ile 350
 340 345

(26) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1116 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- 30 -

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:25:

ATGCCAGGAA ACCGCAACCC AGTGAACACC ACTGCCCCCT GGGCTCCCT GGGCTCTCC 60
 GCCAGACCT GCAACAACT GTCTTCGAA GAGAGCAGAA TACTCTTGT CGTGTGTAC 120
 AGCGGGGTGT GCAAGCTGGG GGTGCCGCC AACTGCTGGA CTGCGTGGCT GGGCTGTCTG 180
 CAGGTACTGC AGGGCAAGT GTTGGCCGTC TACTGTCTT GCTTGGCACT CTGCGACTG 240
 CTGTACAGAG GCAAGCTGCC ACTGTGGGTC ATCTATATCC GCAACAGCA CCGCTGGACC 300
 CTGAGCTTGC TGGCTCTGAA GTTGAACGCC TACTCTTCT TCTGCAAGAT CTAGCTCAGC 360
 ATCTCTTCC TGTGTGCAT CTCTGTGCAC CGCTTGTGG CCGTGTGTA CGCGCTGGAG 420
 AGTGGGGCC GCGCGCCGCG GAGAACGCC ATCTCATCT CGCGCTGCAT CTTCATCTTC 480
 GTCCGGAATG TTCACTACCC GGTGTTCAG ACGAAGACA AGAGACTG CTTCAGATG 540
 CTGCAATGG ACAGCAGAT TGCAGGTAC TACTACGCA GGTTCACCGT TGGCTTGGCC 600
 ATCCCTCTT CCATCATGCC CTTCACACAC CACCGGATTT TCAAGAGAT CAAGCAGAGC 660
 ATGGGCTTAA GCGCTGCCCA GAAAGCCAG GTGAAGACT GCGCATGCG GGTGGTGTTC 720
 ATCTTCTTAA TCTGCTTGC CCGTACACAC CTGGTCTCC TCGTCAGC CCGTCTCTT 780
 TCCTACTACA GAGAGACAG GAACGCCAT TGCAGCTTGG AGAAGAGCT GTACACAGCC 840
 TCTGTGTGT TTCTGTGCT GTCCAGGTG AAGGGGTGG CTGACCCAT TATCTAGCTG 900
 CTGCGCAGCG ACCATCCCG CCAGGAAGTG TCCGAATCC ATAAAGGGTG GAAAGAGTGG 960
 TCCATGAAGA CAGAGCTAC CAGGCTACC CACAGCAGGG ACACCGAGGA GTCTGATGG 1020
 CCGGTGGCCC TTGAGACCA CTACACTTC TCGAGGCCCG TGCACCCACC AGGTACACCA 1080
 TGCCCTGCAA AGAGCTGAT TGAGGAGTCC TGCTGA 1116

(28) INFORMATION FOR SEQ ID NO:26:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 371 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(X1) SEQUENCE DESCRIPTION: SEQ ID NO:26:

Met Pro Gly Asn Ala Thr Pro Val Thr Thr Ala Pro Trp Ala Ser 1 5 10 15 30

- 31 -

Leu Gly Leu Ser Ala Lys Thr Cys Asn Asn Val Ser Phe Gln Gln Ser 20 25 30
 Arg Ile Val Leu Val Val Tyr Ser Ala Val Cys Thr Leu Gly Val 35 40 45
 Pro Ala Asn Cys Leu Thr Ala Trp Leu Ala Leu Leu Gln Val Leu Gln 50 55 60
 Gly Asn Val Leu Ala Val Tyr Leu Leu Cys Leu Ala Leu Cys Gln Leu 65 70 75 80
 Leu Tyr Thr Gly Thr Leu Pro Leu Trp Val Ile Tyr Ile Arg Asn Gln 85 90 95
 His Arg Trp Thr Leu Gly Leu Leu Ala Ser Lys Val Thr Ala Tyr Ile 100 105 110
 Phe Phe Cys Asn Ile Tyr Val Ser Ile Leu Phe Leu Cys Cys Ile Ser 115 120 125
 Cys Asp Arg Phe Val Ala Val Val Tyr Ala Leu Gln Ser Arg Gly Arg 130 135 140
 Arg Arg Arg Arg Thr Ala Ile Leu Ile Ser Ala Cys Ile Phe Ile Leu 145 150 155 160
 Val Gly Ile Val His Tyr Pro Val Phe Gln Thr Gln Asp Lys Gln Thr 165 170 175
 Cys Phe Asp Met Leu Gln Met Asp Ser Arg Ile Ala Gly Tyr Tyr Tyr 180 185 190
 Ala Arg Phe Thr Val Gly Phe Ala Ile Pro Leu Ser Ile Ile Ala Phe 195 200 205
 Thr Asn His Arg Ile Phe Arg Ser Ile Lys Gln Ser Met Gly Leu Ser 210 215 220
 Ala Ala Gln Lys Ala Lys Val Lys His Ser Ala Ile Ala Val Val Val 225 230 235 240
 Ile Phe Leu Val Cys Phe Ala Pro Tyr His Leu Val Leu Val Lys 245 250 255
 Ala Ala Ala Phe Ser Tyr Tyr Arg Gly Asp Arg Asn Ala Met Cys Gly 260 265 270
 Leu Gln Gln Arg Leu Tyr Thr Ala Ser Val Val Phe Leu Cys Leu Ser 275 280 285
 Thr Val Asn Gly Val Ala Asp Pro Ile Ile Tyr Val Leu Ala Thr Asp 290 295 300 35

-32-

His Ser Arg Gln Glu Val Ser Arg Ile His Lys Gly Trp Lys Glu Trp
305 310 315 320
Ser Met Lys Thr Asp Val Thr Arg Leu Thr His Ser Arg Asp Thr Glu
325 330 335
Glu Leu Gln Ser Pro Val Ala Leu Ala Asp His Tyr Thr Phe Ser Arg
340 345 350
Pro Val His Pro Gly Ser Pro Cys Pro Ala Lys Arg Leu Ile Glu
355 360 365

Glu Ser Cys
370

(28) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1113 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATGGCGAACT ATAGCCATGC AGCTGACAAC ATTTTGCAAA ATCTCTGCC TCTACAGCC 60
TTTCTGAAC TGAATCTCTT GGGTTTCATA ATAGAGTCA GCGTGTGGG CAACCTCCTG 120
ATCTCCATTT TGCTAGTGA AGATAAGACC TTGCATAGAG CACCTTACTA CTTCCTGTTG 180
GATCTTTGCT GTTCAGATAT CCTCAGATCT GCAATTTGTT TCCCATTTGT GTTCAACTCT 240
GTCAAAATG GCTCTACCTG GACTTATGG ACTCTGACTT GCAAAGTGAT TGCCTTTCTG 300
GGGGTTTTGT CCGTTTCCA CACTGCTTTC ATGCTCTTCT GCATCATGTGT CACCAGATAC 360
TTAGCTATCG CCCATCACCG CTTCATACA AAGAGGCTGA CTTTTGGAC GTGTCTGGCT 420
GTGATCTGTA TGGTGTGGAC TCTGTCTGTG GCCATGGCAT TTCCCCCGGT TTTAGACGTG 480
GGCATTACT CATTCATTAG GGAGGAAGAT CAATGCACCT TCCAACACCG CTCCTTCAGG 540
GCTAATGANT CCTTAGGANT TATGTGCTT CTGTCTCTCA TCCTCTCTAGC CACACAGCTT 600
GTCTACCTCA AGCTGATATT TTTGCTCCAC GATCGAAGAA AATGAGGCC AGTCCAGTTT 660
GTAGACGAG TCAGCCAGAA CTGGACTTTT CATGCTCCTG GAGCCAGTGG CCAGGCAGCT 720
GCCAATTTGG TAGCAGGANT TGGAGGGGGT CCCACACCAC CCACCTTGCT GGGCATCAGG 780
CAAAATGCAA ACACCACAGG CAGAAGAGG CTATTGGTCT TAGACAGATT CAAATGGAG 840

-33-

AAAAGAATCA GCAGAATGT CTATATAATG ACTTTTCTGT TTCTAACCTT GTGGGCCCC 900
TACTGTGGG CCTGTTATTG GAGAGTTTT GCAGAGGGC CTGTAGTACC AGGGGGATT 960
CTAACAGTG CTGCTGGAT GAGTTTGGC CAAGCAGGA TCAATCCCTT TGCTGCATT 1020
TTCTCACA GAGAGTGAG GCGCTGTTT AGCACACCC TTCTTTACTG CAGAAATCC 1080
5 AGGTTACCAA GGAACCTTA CTGTGTATA TGA 1113

(29) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 370 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Met Ala Asn Tyr Ser His Ala Ala Asp Asn Ile Leu Gln Asn Leu Ser
1 5 10 15
Pro Leu Thr Ala Phe Leu Lys Leu Thr Ser Leu Gly Phe Ile Ile Gly
20 25 30
Val Ser Val Val Gly Asn Leu Leu Ile Ser Ile Leu Leu Val Lys Asp
35 40 45
Lys Thr Leu His Arg Ala Pro Tyr Tyr Phe Leu Leu Asp Leu Cys
50 55 60
Ser Asp Ile Leu Arg Ser Ala Ile Cys Phe Pro Phe Val Phe Asn Ser
65 70 75 80
Val Lys Asn Gly Ser Thr Thr Tyr Gly Thr Leu Thr Cys Lys Val
85 90 95
Ile Ala Phe Leu Gly Val Leu Ser Cys Phe His Thr Ala Phe Met Leu
100 105 110
Phe Cys Ile Ser Val Thr Arg Tyr Leu Ala Ile Ala His His Arg Phe
115 120 125
Tyr Thr Lys Arg Leu Thr Phe Thr Thr Cys Leu Ala Val Ile Cys Met
130 135 140
Val Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Leu Asp Val
145 150 155
Gly Thr Tyr Ser Phe Ile Arg Glu Asp Gln Cys Thr Phe Gln His

- 34 -

165 170 175
 Arg Ser Phe Arg Ala Asn Asp Ser Leu Gly Phe Met Leu Leu Ala
 180 185 190
 Leu Ile Leu Leu Ala Thr Gln Leu Val Tyr Leu Lys Leu Ile Phe Phe
 195 200 205
 Val His Asp Arg Arg Lys Met Lys Pro Val Gln Phe Val Ala Ala Val
 210 215 220
 Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Ser Gly Gln Ala Ala
 225 230 235 240
 Ala Asn Trp Leu Ala Gly Phe Gly Arg Gly Pro Thr Pro Thr Leu
 245 250 255
 Leu Gly Ile Arg Gln Asn Ala Asn Thr Thr Gly Arg Arg Leu Leu
 260 265 270
 Val Leu Asp Gln Phe Lys Met Gln Lys Arg Ile Ser Arg Met Phe Tyr
 275 280 285
 Ile Met Thr Phe Leu Phe Leu Thr Leu Trp Gly Pro Tyr Leu Val Ala
 290 295 300
 Cys Tyr Trp Arg Val Phe Ala Arg Gly Pro Val Val Pro Gly Gly Phe
 305 310 315 320
 Leu Thr Ala Ala Val Trp Met Ser Phe Ala Gln Ala Gly Ile Asn Pro
 325 330 335
 Phe Val Cys Ile Phe Ser Asn Arg Gln Leu Arg Arg Cys Phe Ser Thr
 340 345 350
 Thr Leu Leu Tyr Cys Arg Lys Ser Arg Leu Pro Arg Gln Pro Tyr Cys
 355 360 365
 Val Ile
 370

(30) INFORMATION FOR SEQ ID NO:29:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1080 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:29:
 ATGCAAGTCC GAAACAGCAC CGGCGCGGAC AACGCGAGCG TGCAGTCTCT GCGGAACCGG 60

- 35 -

GCGATGCGCG TGAGCCTGACC CGTGTGTAC TGGCTGTGTG CGGCGGTGAC CATCCCGGGC 120
 AACCTCTCT CTCGTGTGGT GCTGTGCGCG CGCATGAGGAC CCAGATCCCG GTGGGTATTC 180
 TTGATGATCA ACCTGAGCGT CAGGAGCTG ATGTGAGCCA GCGTGTGCG TTTCGAATTC 240
 TACTACCAAT GCACACGCCA CCAGTGGGTA TTGCGGGTGTG TGCTTTGGCA CGTGTGACC 300
 GTGAGCCTTT ACGAACAAT GTATCCAGC ATCTGACCA TGAACGTAT CAGGTGAGG 360
 CGCTTCTG GAGTCTGTGA CCGCTCAGC TCAGAGCGCT GCGCGCGCGG TCGTACGCG 420
 GTGCGCGCGT GTGCAAGGAC CTGACTGTCT CTCCTGACCG CCTGTGTGCC GCTGCGCGC 480
 ACCGATCTCA CCTACCGCGT GCAGGCCCTG GGCATGATCA CTTGCTTGA CGTCTCAAG 540
 TGGACGATGC TCCCAAGCGT GGCATGTGG GCGGTGTTC TTTCACCAT CTTCATCTG 600
 CTGTTCTCA TCCGTTCTGT GATCACCCTG GCTTGTACA CGGCAACCAT CTTCAACTG 660
 TTGCGCACCG AGAGAGCGCA CGGCGGGGAG CAGCGAGGC GCGCGGTGGG CTTGCGCGG 720
 GTGTCTTGC TGACCTTGT CACTGTCTG GCCCCACACA ACTTGTTGCT CTTGCGCGC 780
 ATCGTAGCC GCGTGTCTGA CGGCAAGAG TACTACAGG TGTACAAGCT CAGCGTGTG 840
 CTCAGCTCC TCACACATG TGTGACCGG TTGTATT ACTTGTGCT CCGGAAATC 900
 CAGCTGCGCC TCGGGAATA TTGAGGCTGC CGCGGGTGC CCAGACAC CTTGACAGC 960
 CGCCGAGGA GCCTTCTCT CGCAGAGAC AGTCCGTGCT GTTCGAGGC CGGTGCGAC 1020
 CTTGAAGGA TGGAGGAGAC CACAGAGCC GGCCTCCAGA GGCAGAGAG TGTGTTCTGA 1080

(31) INFORMATION FOR SEQ ID NO:30:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 359 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:
 Met Gln Val Pro Asn Ser Thr Gly Pro Asp Asn Ala Thr Leu Gln Met 15
 Leu Arg Asn Pro Ala Ile Ala Val Ala Leu Pro Val Val Tyr Ser Leu 20
 Val Ala Ala Val Ser Ile Pro Gly Asn Leu Phe Ser Leu Trp Val Leu 25

-36-

35 40 45
Cys Arg Arg Met Gly Pro Arg Ser Pro Ser Val Ile Phe Met Ile Asn 60
50
Leu Ser Val Thr Asp Leu Met Leu Ala Ser Val Leu Pro Phe Gln Ile 80
65 70
Tyr Tyr His Cys Asn Arg His His Trp Val Phe Gly Val Leu Leu Cys 95
85
Asn Val Val Thr Val Ala Phe Tyr Ala Asn Met Tyr Ser Ser Ile Leu 110
100
Thr Met Thr Cys Ile Ser Val Glu Arg Phe Leu Gly Val Leu Tyr Pro 115
120
Leu Ser Ser Lys Arg Trp Arg Arg Arg Tyr Ala Val Ala Ala Cys 130
135
Ala Gly Thr Trp Leu Leu Leu Thr Ala Leu Cys Pro Leu Ala Arg 145
150 155 160
Thr Asp Leu Thr Tyr Pro Val His Ala Leu Gly Ile Ile Thr Cys Phe 165
170 175
Asp Val Leu Lys Trp Thr Met Leu Pro Ser Val Ala Met Trp Ala Val 180
185 190
Phe Leu Phe Thr Ile Phe Ile Leu Leu Phe Leu Ile Pro Phe Val Ile 195
200 205
Thr Val Ala Cys Tyr Thr Ala Thr Ile Leu Lys Leu Leu Arg Thr Glu 210
215 220
Glu Ala His Gly Arg Glu Gln Arg Arg Ala Val Gly Leu Ala Ala 225
230 235 240
Val Val Leu Leu Ala Phe Val Thr Cys Phe Ala Pro Asn Asn Phe Val 245
250 255
Leu Leu Ala His Ile Val Ser Arg Leu Phe Tyr Gly Lys Ser Tyr Tyr 260
265 270
His Val Tyr Lys Leu Thr Leu Cys Leu Ser Cys Leu Asn Asn Cys Leu 275
280 285
Asp Pro Phe Val Tyr Tyr Phe Ala Ser Arg Glu Phe Gln Leu Arg Leu 290
295 300
Arg Glu Tyr Leu Gly Cys Arg Arg Val Pro Arg Asp Thr Leu Asp Thr 305
310 315 320
Arg Arg Glu Ser Leu Phe Ser Ala Arg Thr Thr Ser Val Arg Ser Glu 325
330 335

-37-

Ala Gly Ala His Pro Glu Gly Met Glu Gly Ala Thr Arg Pro Gly Leu 340
345 350
Gln Arg Gln Glu Ser Val Phe 355
5 (32) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1503 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
10
(ii) MOLECULE TYPE: DNA (genomic)
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:
ATGAGCGTTC CTGCGGAGGA CAGCCAGGC CCGAGGGGG CAGCTAGGG CTGCGCTGTG 60
CCAGTCGCG CCGGGGGCG CTCGGTGCC GCGGGAGTG GCACAGGCTG GCAGCCATGG 120
15 GCTGAGTGCC CCGGACCCAA GGGGAGGGG CAACTGCTGG CGACCCCGG CCCTTTGCGT 180
CGCTGCGCG CCCCTCTGCC TGCCAGTCC AGCCCGGCC CCGAGCGGC GTCCGCTCAC 240
TCGCTTCAAG GCAGCGGAC TCGGGTGGC GCACGACCAG GCGCGAGACC TTGGGGCGCG 300
CGGCCCATGG AGTGGGGGCT GCTGCGGCG GCGCCGGTGA GCGAGTCAAT CGTCTGTCAT 360
TACACTACA CCGGCAAGCT CCGGGTGGC AGCTACCAG CCGGTGCGG CCTGCGCGCC 420
20 GACGCGTGG TGTGCTTGG GGTGTGCGC TTCATCGTGC TAGAGATCT AGCCGTGTTG 480
TTGTTGCTGG GAGCCACCC GCGCTTCCAC GCTCCCATGT TCCTGCTCT GGGCAGCCTC 540
ACGTTGTGG ATCTGTGAG AGGCGCGCC TAGCGCGCCA ACATCTACT GTCGGGGCG 600
CTCAGCTGA AACTGTCCCC CGGCTCTGG TTGCGACGG AGGAGGCGT CTTCTGTGCA 720
CTCAGTGGT CCGTGTGAG CCTCTTGGC ATCGGCTGG AGCGCAGCT CACCATGGCG 780
25 CCGAGGGGG CCGGCGCCGT CTCAGTGGG GGGCGCAGC TGGCGATGGC AGCCGCGGCC 840
TGGGGGCTG CGCTGTCTCT CGGGCTCTG CCAGCGCTGG GCTGGATTTG CTTGGGTGCG 900
CTGAGGCTT GCTCCACTGT CTTGCGCTC TAGCGCAAG CTTACGTGT CTTCTGCTG 960
CTGCGCTTG TGGGATCTT GCGCGGATC TGTGACTCT ACAGCGCAT CTACTGCCAG 1020
GTAGCGCCA ACGGCGGGG CTTGCCGCA CCGCCCGGGA CTGCGGGGAC CACTCGACC 1080
30 CCGGCGCTG GCAAGCGCG CTTCTTGGC TTGCTGCGCA CGCTCAGCT GGTGCTCTG 1080

- 38 -

GCCTTGTG CATGTGGGG CCCCCCTTC CTGCTGCTGT TGCTGACGT GGGGTGGCCG 1140
 GGGGACACT GTCCCTACT CCTGAGAGCC GATCCCTTCC TGGGACTGAGC CATGGCCAAC 1200
 TCACCTCTGA ACCCCATCAT CTACAGCCTC ACCAACCGCG ACCTGGCGCA CGCGCTCTCG 1260
 CGCTGGGTCT GGTGGGAGAC CCACTCTTGC GAGCAGAGAC CAGTGGCTC CAGCAGTGG 1320
 GCGAGCGCG CTGAGGCTTC CGGGGGCTTG CGCGCTTCCC TGCCCCCGGG CTTGATGGG 1380
 ACCTTCAGCG GCTCGAGAC CTCATCGGCC CAGCGGAGCG GCGTGGACAC CAGCGGCTCC 1440
 ACAGGCAACC CGGTGTCAC CACAGCCGCC CGGACTCTGG TATCAGAAC GCGTGGAGAC 1500
 TGA 1503

(33) INFORMATION FOR SEQ ID NO:32:

10 (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 500 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

15 (11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Met Glu Arg Pro Trp Glu Asp Ser Pro Gly Pro Glu Gly Ala Ala Glu 15
 1 5 10
 Gly Ser Pro Val Pro Val Ala Ala Gly Ala Arg Ser Gly Ala Ala 25
 20 25 30
 Ser Gly Thr Gly Trp Gln Pro Trp Ala Glu Cys Pro Gly Pro Lys Gly 45
 35 40 45
 Arg Gly Gln Leu Leu Ala Thr Ala Gly Pro Leu Arg Arg Trp Pro Ala 60
 50 55 60
 Pro Ser Pro Ala Ser Ser Ser Pro Ala Pro Gly Ala Ala Ser Ala His 80
 65 70 75
 Ser Val Gln Gly Ser Ala Thr Ala Gly Gly Ala Arg Pro Gly Arg Arg 95
 85 90 95
 Pro Trp Gly Ala Arg Pro Met Glu Ser Gly Leu Leu Arg Pro Ala Pro 110
 100 105 110
 Val Ser Glu Val Ile Val Leu His Tyr Asn Tyr Thr Gly Lys Leu Arg 125
 115 120 125
 Gly Ala Ser Tyr Gln Pro Gly Ala Gly Leu Arg Ala Asp Ala Val Val 140
 130 135 140

- 39 -

Cys Leu Ala Val Cys Ala Phe Ile Val Leu Glu Asn Leu Ala Val Leu 160
 145 150 155
 Leu Val Leu Gly Arg His Pro Arg Phe His Ala Pro Met Phe Leu Leu 175
 165 170
 Leu Gly Ser Leu Thr Leu Ser Asp Leu Leu Ala Gly Ala Tyr Ala 190
 180 185
 Ala Asn Ile Leu Leu Ser Gly Pro Leu Thr Leu Lys Leu Ser Pro Ala 205
 195 200
 Leu Trp Phe Ala Arg Glu Gly Val Phe Val Ala Leu Thr Ala Ser 220
 210 215
 Val Leu Ser Leu Leu Ala Ile Ala Leu Glu Arg Ser Leu Thr Met Ala 240
 225 230 235
 Arg Arg Gly Pro Ala Pro Val Ser Ser Arg Gly Arg Thr Leu Ala Met 255
 245 250
 Ala Ala Ala Tyr Gly Val Ser Leu Leu Leu Gly Leu Leu Pro Ala 270
 260 265
 Leu Gly Trp Asn Cys Leu Gly Arg Leu Asp Ala Cys Ser Thr Val Leu 285
 275 280
 Pro Leu Tyr Ala Lys Ala Tyr Val Leu Phe Cys Val Leu Ala Phe Val 300
 290 295
 Gly Ile Leu Ala Ala Ile Cys Ala Leu Tyr Tyr Ile Tyr Cys Gln 320
 305 310 315
 Val Arg Ala Asn Ala Arg Arg Leu Pro Ala Arg Pro Gly Thr Ala Gly 335
 325 330
 Thr Thr Ser Thr Arg Ala Arg Arg Lys Pro Arg Ser Leu Ala Leu Leu 350
 340 345
 Arg Thr Leu Ser Val Val Leu Leu Ala Phe Val Ala Cys Trp Gly Pro 365
 355 360
 Leu Phe Leu Leu Leu Leu Leu Asp Val Ala Cys Pro Ala Arg Thr Cys 380
 370 375
 Pro Val Leu Leu Gln Ala Asp Pro Phe Leu Gly Leu Ala Met Ala Asn 400
 385 390 395
 Ser Leu Leu Asn Pro Ile Ile Tyr Thr Leu Thr Asn Arg Asp Leu Arg 415
 405 410
 His Ala Leu Leu Arg Leu Val Cys Cys Gly Arg His Ser Cys Gly Arg 430
 420 425
 Asp Pro Ser Gly Ser Gln Gln Ser Ala Ser Ala Ala Glu Ala Ser Gly 440

- 40 -

435 440 445
 Gly Leu Arg Arg Cys Leu Pro Gly Leu Asp Gly Ser Phe Ser Gly
 450 455 460
 Ser Glu Arg Ser Ser Pro Gln Arg Asp Gly Leu Asp Thr Ser Gly Ser
 465 470 475 480
 Thr Gly Ser Pro Gly Ala Pro Thr Ala Ala Arg Thr Leu Val Ser Glu
 485 490 495
 Pro Ala Ala Asp
 500

10 (34) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1029 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ATGCAAGCG TCGACATCT CACCTCTGG CCTGGACCA CCACTGTGT CACCAGAGAC 60
 TACAAATCA CCCAGTTCCT CTTCCACTG CTCTACACTG TCCTGTTTTT TGTGGACTT 120
 20 ATCACAATG GCTGGCGAT GAGGATTTT TTTCAATCC GAGTAATATC AAACCTTATT 180
 ATTTTCTTA AGAACACAGT CATTCTGAT CTCTCATGA TTCTGACTTT TCCATTCAAA 240
 ATTCTTAGTG ATGCCAACT GGGAACAGGA CCACTGAGAA CTTTGTGTG TCAAGTTACC 300
 TCCGTCAAT TTTATTTTAC ATGTATATC AGTATTTCAT TCCTGGGACT GATAACTATC 360
 GATCGTACC AGAAGACCC CAGGCCATT AAAACATCCA ACCCAAAA TCCTTGGGG 420
 25 GCTAAGATT TCCTGTGTG CATCTGGCA TTCAATGTT TACTCTCTT GCCTAACATG 480
 ATTCTGACCA ACAGGCAGCC GAGAGACAAG ANTGTGAGA ANTGTCTTT CTTAAATCA 540
 GAGTTCGGTC TAGTCTGGCA TGAATAGTA AATTACATCT GTCAAGTCAT TTCTGGATT 600
 AATTCTTAA TTGTTATCT ATGTTATCA CTCAATACAA AAGAAGTGA CCGGTATAC 660
 GTAAGAACA GGGGTGTAGG TAAAGTCCC AGGAABAGG TGAAGTCGA AGTTTCATT 720
 30 ATCATTTGCT TATCTTTAT TTGTTTGTG CTTTCCATT TTGCCCGAAT TCCTTACACC 780
 CTGAGCAAA CCGGGATGT CTTTGTCTG ACTGTGAAA ATACTCTGTT CTATGTGAAA 840

- 41 -

GAGAGCACT TGTGTTTAC TTCTTAAT GCATGCTGG ATCCGTTGAT CTATTTTTTC 900
 CTTTGCAGT CCTTCAGAA TTCTTTGATA AGTATGCTGA AGTGCCCAA TTTCGCAACA 960
 TCTCTGCC AGGACAATAG GAAAAAGAA CAGGATGCTG GTGACCCAAA TGAAGAGACT 1020
 CCAATGTAA 1029

5 (35) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 342 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Met Gln Ala Val Asp Asn Leu Thr Ser Ala Pro Gly Asn Thr Ser Leu 15
 1 5 10
 Cys Thr Arg Asp Tyr Lys Ile Thr Gln Val Leu Phe Pro Leu Leu Tyr 30
 20 25
 Thr Val Leu Phe Phe Val Gly Leu Ile Thr Asn Gly Leu Ala Met Arg 45
 35 40
 Ile Phe Phe Gln Ile Arg Ser Lys Ser Asn Phe Ile Ile Phe Leu Lys 60
 50 55
 Asn Thr Val Ile Ser Asp Leu Leu Met Ile Leu Thr Phe Pro Phe Lys 80
 65 70 75
 Ile Leu Ser Asp Ala Lys Leu Gly Thr Gly Pro Leu Arg Thr Phe Val 95
 85 90
 Cys Gln Val Thr Ser Val Ile Phe Tyr Phe Thr Met Tyr Ile Ser Ile 110
 100 105
 Ser Phe Leu Gly Leu Ile Thr Ile Asp Arg Tyr Gln Lys Thr Thr Arg 125
 115 120
 Pro Phe Lys Thr Ser Asn Pro Lys Asn Leu Leu Gly Ala Lys Ile Leu 140
 130 135
 Ser Val Val Ile Trp Ala Phe Met Phe Leu Leu Ser Leu Pro Asn Met 160
 145 150 155
 Ile Leu Thr Asn Arg Gln Pro Arg Asp Lys Asn Val Lys Lys Cys Ser 175
 165 170
 Phe Leu Lys Ser Glu Phe Gly Leu Val Trp His Glu Ile Val Asn Tyr 35

-42-

180 185 190
 Ile Cys Gln Val Ile Phe Trp Ile Asn Phe Leu Ile Val Ile Cys
 195 200 205
 Tyr Thr Leu Ile Thr Lys Glu Leu Tyr Arg Ser Tyr Val Arg Thr Arg
 210 215 220
 Gly Val Gly Lys Val Pro Arg Lys Lys Val Asn Val Lys Val Phe Ile
 225 230 235 240
 Ile Ile Ala Val Phe Phe Ile Cys Phe Val Pro Phe His Phe Ala Arg
 245 250 255
 Ile Pro Tyr Thr Leu Ser Gln Thr Arg Asp Val Phe Asp Cys Thr Ala
 260 265 270
 Glu Asn Thr Leu Phe Tyr Val Lys Glu Ser Thr Leu Trp Leu Thr Ser
 275 280 285
 Leu Asn Ala Cys Leu Asp Pro Phe Ile Tyr Phe Phe Leu Cys Lys Ser
 290 295 300
 Phe Arg Asn Ser Leu Ile Ser Met Leu Lys Cys Pro Asn Ser Ala Thr
 305 310 315 320
 Ser Leu Ser Gln Asp Asn Arg Lys Lys Glu Gln Asp Gly Gly Asp Pro
 325 330 335
 Asn Glu Glu Thr Pro Met
 340

(36) INFORMATION FOR SEQ ID NO:35:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1077 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:35:

30 ATGTGAGTCT GCTACGCTCC CCCAGGGAAC GAGACACTGC TGAAGTGGAA GACTTCCGCG 60
 GCCACAGGCA CAGCCTTCTT GCTGCTGGCG GCGTCTCTGG GGTGCTCTGG CAACGGCTTC 120
 GTGGTGTGA GCTTGGCGGG CTGGCGGCTT GCACGGGGGG GACCGCTGAC GGCACAGCTT 180
 GTGCTGCACC TGGCGCTGGC CGACGGCGCG GTGCTGCTGC TCAAGCGGCT CTTTGTGGCC 240
 TTCTGAGACC GGCAGGCTGG GCGCTGGGC CAGGCGGGCT GCAAGCGGCT GTACTACGTG 300

-43-

TGGCGCTGCA GCATGTAGCC CAGCGTGTG CTCACCGGAC TGTCTAGCCT GCACGCTGAC 360
 CTGCGAGTCA CCCGCCCTT CTGAGGAGCT CGGCTGAGCA GCGCGGCGCT GAGCCGCGCG 420
 CTGCTGCTGG CGGTCTGGCT GCGCGGCTGG TGGTGTGCGG TCCCGGCGCG CGTCTACCGC 480
 CACTGTGGA GGAACCGGCT ATGCAGCTG TGCACCGCT CCGCGGTCA CGCGCGCGCC 540
 CACTGAGCC TGAAGACTT GACCGCTTC GTGCTTCTT TGGGGCTGAT GCTCGGCTGC 600
 TACAGCGTGA CGCTGGCAGG GCTGCGGGGG GCGCGCTGGG GCTCGGGGCG GCACGGGGCG 660
 CGGTTGGGCC GCGTGTGAG CAGCATGCTG CTGCTCTGG GCTTGTCTTG GCGCCCTTAC 720
 CAGGAGTCA ACCTTCTGCA GCGGTGCA GCGCTGGCTC CACCGGAGG GGCCTTGGCG 780
 AAGCTGGCG GAGCGGCGCA GCGGCGCGCA GCGGAGACTA CGGCTTGGCG CTTCCTTCACT 840
 TCTAGGCTCA ACCGGTCT CTACGCTTC ACCGTGGAG ATCTGTGCG CCGGGCAGGT 900
 CCCGTTCC TCACGGGCT CTTCGAGAGG TCTGGGGAGG CCGAGGGGG CGGCGGCTCT 960
 AGGAGAGGA CCATGAGCT CCGACTACC CTCAGCTGA AAGTGTGGG GCGAGGCGCG 1020
 GGCATGAGG ACCCGGGGG TGGATGAG AGAGACGGTC CGAATGGA CTTTGA 1077

(37) INFORMATION FOR SEQ ID NO:36:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 358 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:36:

1 Met Ser Val Cys Tyr Arg Pro Gly Asn Glu Thr Leu Leu Ser Trp 15
 5
 1 Lys Thr Ser Arg Ala Thr Gly Thr Ala Phe Leu Leu Leu Ala Leu 25
 20
 25 Leu Gly Leu Pro Gly Asn Gly Phe Val Val Trp Ser Leu Ala Gly Trp 30
 35
 30 Arg Pro Ala Arg Gly Arg Pro Leu Ala Ala Thr Leu Val Leu His Leu 45
 50
 65 Ala Leu Ala Asp Gly Ala Val Leu Leu Leu Thr Pro Leu Phe Val Ala 60
 70
 75 Phe Leu Thr Arg Gln Ala Trp Pro Leu Gly Gln Ala Gly Cys Lys Ala 80

- 44 -

85 90 95
Val Tyr Tyr Val Cys Ala Leu Ser Met Tyr Ala Ser Val Leu Leu Thr
100 105 110
Gly Leu Leu Ser Leu Gln Arg Cys Leu Ala Val Thr Arg Pro Phe Leu
115 120 125
Ala Pro Arg Leu Arg Ser Pro Ala Leu Ala Arg Arg Leu Leu Leu Ala
130 135 140
Val Trp Leu Ala Ala Leu Leu Leu Ala Val Pro Ala Ala Val Tyr Arg
145 150 155 160
His Leu Trp Arg Asp Arg Val Cys Gln Leu Cys His Pro Ser Pro Val
165 170 175
His Ala Ala His Leu Ser Leu Glu Thr Leu Thr Ala Phe Val Leu
180 185 190
Pro Phe Gly Leu Met Leu Gly Cys Tyr Ser Val Thr Leu Ala Arg Leu
195 200 205
Arg Gly Ala Arg Trp Gly Ser Gly Arg His Gly Ala Arg Val Gly Arg
210 215 220
Leu Val Ser Ala Ile Val Leu Ala Phe Gly Leu Leu Trp Ala Pro Tyr
225 230 235 240
His Ala Val Asn Leu Leu Gln Ala Val Ala Ala Leu Ala Pro Pro Glu
245 250 255
Gly Ala Leu Ala Lys Leu Gly Ala Gly Gln Ala Ala Arg Ala Gly
260 265 270
Thr Thr Ala Leu Ala Phe Phe Ser Ser Val Asn Pro Val Leu Tyr
275 280 285
Val Phe Thr Ala Gly Asp Leu Leu Pro Arg Ala Gly Pro Arg Phe Leu
290 295 300
Thr Arg Leu Phe Glu Gly Ser Gly Glu Ala Arg Gly Gly Arg Ser
305 310 315 320
Arg Glu Gly Thr Met Glu Leu Arg Thr Thr Pro Gln Leu Lys Val Val
325 330 335
Gly Gln Gly Arg Gly Asn Gly Asp Pro Gly Gly Gly Met Glu Lys Asp
340 345 350
Gly Pro Glu Trp Asp Leu
355

(38) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 334 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein

WO 00/22131

PCT/US99/24065

- 45 -

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1005 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ATGCTGGGGA TCATGGCATG GAATGCAACT TGCAAAACCT GGCTGGCAGC AGAGGCTGCC 60
 CTGGAAAAGT ACTACCTTTC CATTTTTAT GGGATTGAGT TCGTTGTGGG AGTCCTTGA
 10 AATACCATG TTGTTACGG CTACATCTTC TCCTGAAGA ACTGGACAG CAGTAATATT
 TATCTCTTA ACTCTCTGT CTCTGACTTA GCTTTCTGT GCACCTCCCC CATGCTGATA
 AGGAGTTAT CCAATGAAA CTGGATATAT GGAGACGTGC TGTGCATAAG CAACCGATAT
 20 GTGCTCATG CCAACCTCTA TACCAGCATT CTCCTTCTCA CTTTATCAG CATAGATCGA
 TACTTGATAA TTAAGTATCC TTTCGAGAA CACCTTCTGC AAAAAGAA GTTTGTCTATT
 30 TTAATCTCT TGGCCATTG GGTTTAGTA ACCTTAGAT TACTACCAT ACTTCCCTTT
 40 ATAAATCTG TTAATCTGA CAAGGCACC ACCTGTAATG ATTTGCAAG TTCTGGAGAC
 CCAACTACA ACCTCATTTA CAGCATGTGT CTAACACTGT TGGGGTTCTT TATTCTCTTT
 50 TTGTGATGT GTTCTTTTA TTACAAGATT GCTCTCTTCC TAAAGCAGAG GAATAGGCAG
 60 GTTGCTACTG CTCGCCCCCT TGAAGACCT CTCAACTTGG TCATCATGSC AGTGGTAATC
 70 TTCTCTGTGC TTTTACACC CTATCAGTTC ATGCGAATG TGAGGATGCG TTCACGCTTG
 80 GGGAGTTGA AGCAGTATCA GTGCACTCAG GTGCTCATCA ACTCTTTTTC CATTTGTGACA
 90 CGGCTTTGG CTTTCTGAA CAGTGTATC AACCTGTCT TCTATTTTCT TTGGGAGAT
 CACTTCAGG ACATGCTGAT GAATCAACTG AGACACAAC TCAATCCCT TACATCTTTT
 1005 AGCAGATGG CTCATGAAC TCTACTTCA TTCAGAGAAA AGTGA

(39) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 334 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
- (ii) MOLECULE TYPE: protein

- 46 -

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Met Leu Gly Ile Met Ala Trp Asn Ala Thr Cys Lys Asn Trp Leu Ala
 1 5 10 15
 Ala Glu Ala Ala Leu Glu Lys Tyr Tyr Leu Ser Ile Phe Tyr Gly Ile
 20 25 30
 Glu Phe Val Val Gly Val Leu Gly Asn Thr Ile Val Val Tyr Gly Tyr
 35 40 45
 Ile Phe Ser Leu Lys Asn Trp Asn Ser Ser Asn Ile Tyr Leu Phe Asn
 50 55 60
 Leu Ser Val Ser Asp Leu Ala Phe Leu Cys Thr Leu Pro Met Leu Ile
 65 70 75 80
 Arg Ser Tyr Ala Asn Gly Asn Trp Ile Tyr Gly Asp Val Leu Cys Ile
 85 90 95
 Ser Asn Arg Tyr Val Leu His Ala Asn Leu Tyr Thr Ser Ile Leu Phe
 100 105 110
 Leu Thr Phe Ile Ser Ile Asp Arg Tyr Leu Ile Ile Lys Tyr Pro Phe
 115 120 125
 Arg Glu His Leu Leu Glu Lys Lys Glu Phe Ala Ile Leu Ile Ser Leu
 130 135 140
 Ala Ile Trp Val Leu Val Thr Leu Glu Leu Leu Pro Ile Leu Pro Leu
 145 150 155 160
 Ile Asn Pro Val Ile Thr Asp Asn Gly Thr Tyr Cys Asn Asp Phe Ala
 165 170 175
 Ser Ser Gly Asp Pro Asn Tyr Asn Leu Ile Tyr Ser Met Cys Leu Thr
 180 185 190
 Leu Leu Gly Phe Leu Ile Pro Leu Phe Val Met Cys Phe Phe Tyr Tyr
 195 200 205
 Lys Ile Ala Leu Phe Leu Lys Glu Arg Asn Arg Glu Val Ala Thr Ala
 210 215 220
 Leu Pro Leu Glu Lys Pro Leu Asn Leu Val Ile Met Ala Val Val Ile
 225 230 235 240
 Phe Ser Val Leu Phe Thr Pro Tyr His Val Met Arg Asn Val Arg Ile
 245 250 255
 Ala Ser Arg Leu Gly Ser Tyr Lys Glu Tyr Glu Cys Thr Glu Val Val
 260 265 270
 Ile Asn Ser Phe Tyr Ile Val Thr Arg Pro Leu Ala Phe Leu Asn Ser

- 47 -

275 280 285
 Val Ile Asn Pro Val Phe Tyr Phe Leu Leu Gly Asp His Phe Arg Asp
 290 295 300
 Met Leu Met Asn Glu Leu Arg His Asn Phe Lys Ser Leu Thr Ser Phe
 305 310 315 320
 Ser Arg Trp Ala His Glu Leu Leu Leu Ser Phe Arg Glu Lys
 325 330

(40) INFORMATION FOR SEQ ID NO:39:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1296 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:39:

ATGCAGGCGC TTAACTATAC CCGGAGCAG TTCTCTGCGC TGCTGGGGA CCACAACCTG
 60
 AGCGGAGC AGTTCATCGC TCTGACCGC CTGGACGCGC TCGCTACAC CCCAGAGCTG
 120
 CCGGAGCAG CCAAGCTGCG CCTGCTGCTC ACCGGGAGCG TCATCTTGGC CCGGGGCTC
 180
 TTGGGAGTGT CTCTGATGTT CTACGCTGTT ACCCGAGCA AGGCAATGCG CACCGTACCC
 240
 AACATCTTTA TCTGCTCTCT GCGGCTCAGT GACCTGCTCA TCACCTTCTT CTGCAATGCC
 300
 GTCAACATGC TCCAGAACAT TTCCGACAC TGCTGGGCG GTGCTTTCA TTGCAGAGATG
 360
 GTGCCATTTG TCCAGTCTAC CGCTGTTG AGAGAAATGC TCATCTATGAC CTGCATTGCT
 420
 GTGAAAGGC ACCAGGAGCT TGTGATCCT TTAAATGA AGTGCATTA CACCAACGA
 480
 AGGCTTTCA CAATGCTAGG TGTGCTGCG CTGGTGGGAG TCATGCTAGG ATCACCCTAG
 540
 TGGCAGGTGC AACAACTTGA GATCAATAT GACTTCTTAT ATGAAAGGA ACACATCTGC
 600
 TGCTTAGAG AGTGCACGAG CCTGTGCG CAGAAATCT ACACACCTT CATCTGTGTC
 660
 ATCTCTTCC TCTGCTCTCT TATGTGATG CTATCTCTGT ACAGTAAT TTGTTATGAA
 720
 CTGTGATTA AGAAAAGAT TGGGATAGGT TCAGTGCTTC GAATATTA TGAAAAGAA
 780
 ATGTCCAAA TAGCCAGGA GAAGAAAGA GCTGTCAATA TGATGTGAC AGTGTGGCT
 840
 CTCTTGTCTG TGTGCTGGGC ACCATTCAT GTTGTCAATA TGATGATTA ATACAGTAAT
 900
 TTGAAAGG AATATATGA TGTCAATC AGATGATTT TTGCTATGCT GCAATTAAT
 960

GGATTTCCTCA ACTCCACTG TAATCCCAAT GTCTATGCAAT TTATGAATGA AAATCTCAAA 1020
 AAAAATGTTT TGTCTGCAGT TTGTTATTGC ATAGTAATAA AAACCTTCTC TCACGACAAA 1080
 AGGCATGGAA ATTCCAGGAT TACAATGATG CGGAGAGAG CAAAGTTTTC CCTCAGAGAG 1140
 AATCCAGTGG AGGAAACCAA AGGAGAGACA TTCAGTGAAG GCACATTTGA AGTCAAAATTG 1200
 5 TGTGAACAGA CAGAGGAGAA GAAAGAGCTC AAACGACATC TTGCTCTCTT TAGGTCTGAA 1260
 CTGCTTGAGA ATTCTCTTT AGACAGTGGG CATTAA 1296

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 431 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

15 Met Gln Ala Leu Asn Ile Thr Pro Glu Gln Phe Ser Arg Leu Leu Arg 15
 1 AAAATGTTT TGTCTGCAGT TTGTTATTGC ATAGTAATAA AAACCTTCTC TCACGACAAA 1080
 Asp His Asn Leu Thr Arg Glu Gln Phe Ile Ala Leu Tyr Arg Leu Arg 30
 20 Pro Leu Val Tyr Thr Pro Glu Leu Pro Gly Arg Ala Lys Leu Ala Leu 45
 35 Val Leu Thr Gly Val Leu Ile Phe Ala Leu Ala Leu Phe Gly Asn Ala 60
 50 Leu Val Phe Tyr Val Val Thr Arg Ser Lys Ala Met Arg Thr Val Thr 80
 65 Asn Ile Phe Ile Cys Ser Leu Ala Leu Ser Asp Leu Leu Ile Thr Phe 95
 85 Phe Cys Ile Pro Val Thr Met Leu Gln Asn Ile Ser Asp Asn Trp Leu 110
 100 Gly Gly Ala Phe Ile Cys Lys Met Val Pro Phe Val Gln Ser Thr Ala 125
 115 Val Val Thr Glu Met Leu Thr Met Thr Cys Ile Ala Val Glu Arg His 140
 130 Gln Gly Leu Val His Pro Phe Lys Met Lys Trp Gln Tyr Thr Asn Arg 155
 145

Arg Ala Phe Thr Met Leu Gly Val Val Trp Leu Val Ala Val Ile Val 175
 165 Gly Ser Pro Met Trp His Val Gln Gln Leu Glu Ile Lys Tyr Asp Phe 190
 180 Leu Tyr Glu Lys Glu His Ile Cys Cys Leu Glu Glu Trp Thr Ser Pro 205
 195 Val His Gln Lys Ile Tyr Thr Thr Phe Ile Leu Val Ile Leu Phe Leu 220
 210 Leu Pro Leu Met Val Met Leu Ile Leu Tyr Ser Lys Ile Gly Tyr Glu 240
 225 Leu Trp Ile Lys Lys Arg Val Gly Asp Gly Ser Val Leu Arg Thr Ile 255
 245 His Gly Lys Glu Met Ser Lys Ile Ala Arg Lys Lys Lys Arg Ala Val 270
 260 Ile Met Met Val Thr Val Val Ala Leu Phe Ala Val Cys Trp Ala Pro 285
 275 Phe His Val Val His Met Met Ile Glu Tyr Ser Asn Phe Glu Lys Glu 300
 295 Tyr Asp Asp Val Thr Ile Lys Met Ile Phe Ala Ile Val Gln Ile Ile 320
 305 Gly Phe Ser Asn Ser Ile Cys Asn Pro Ile Val Tyr Ala Phe Met Asn 335
 325 Glu Asn Phe Lys Lys Asn Val Leu Ser Ala Val Cys Tyr Cys Ile Val 350
 340 Asn Lys Thr Phe Ser Pro Ala Gln Arg His Gly Asn Ser Gly Ile Thr 365
 355 Met Met Arg Lys Lys Ala Lys Phe Ser Leu Arg Glu Asn Pro Val Glu 380
 370 Glu Thr Lys Gly Glu Ala Phe Ser Asp Gly Asn Ile Glu Val Lys Leu 400
 385 Cys Glu Gln Thr Glu Glu Lys Lys Lys Leu Lys Arg His Leu Ala Leu 415
 405 Phe Arg Ser Glu Leu Ala Glu Asn Ser Pro Leu Asp Ser Gly His 430
 420

35 (42) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 24 base pairs

-50-

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CTGTGTACAG CAGTTGCGAG AGTG

24

(43) INFORMATION FOR SEQ ID NO:42:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:42:

15 GAGTGCAGG CAGACGAGT AGAC

24

(44) INFORMATION FOR SEQ ID NO:43:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:43:

25 CCCGAAATTC TCGTTGCTCC CAGCTTGACC C

31

(45) INFORMATION FOR SEQ ID NO:44:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: YES

-51-

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:44:
TGTGATCTCT GCTGTCAAAG GCCCAATTC GG

32

(46) INFORMATION FOR SEQ ID NO:45:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

10 (1v) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:45:

TCACAACTCT AGGTGTGATC

20

(47) INFORMATION FOR SEQ ID NO:46:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

20 (1v) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TGCATAGACA ATGGATTAC AG

22

(48) INFORMATION FOR SEQ ID NO:47:

- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 511 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TCACAACTCT AGGTGTGATC TCGTGTGTG CAGTCATCT AGATCAACC ATGTGCGACG
TGCACAACT TGAGATCAAA TATGACTTCC TATATGAAAA GGAACACATC TCGTGTGTG

60

120

- 52 -

AAGAGTGAC CAGCCCTGTG CACGAGAGA TCTACACCAC CTTATCCTT GTCATCTCT 180
 TCTCTCTGCC TCTTATGGTG ATGCTTATTG TGTACGTAA ATTGGTTATG AACTTTGGAT 240
 AAAGAAAAGA GTTGGGGATG GTTCACTGCT TCGAACTATT CATTGAAAAG AAATGTCCAA 300
 AATAGCCAGG AAGAGAAAC GAGCTGTCTAT TATGATGGTG ACAGTGGTGG CTCTCTTTGC 360
 5 TGTGTGCTGG GCACCAATCC ATGTGTCCA TATGATGATT GAATACAGTA ATTTGAAAA 420
 GGAATATGAT GATGTACAA TCAAGATGAT TTTTGCTATC GTGCAAAATTA TTGGATTTTC 480
 CCACTCCATC TGTATCCCA TTGTCTATGC A 511
 (49) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CTGCTTAGAA GAGTGACCA G

(50) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25 (iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CTGTGCACCA GAAGATCTAC AC

(51) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- 53 -

(ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: YES
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:
 CAAGGATGAA GGTTGGTGTAG A

5 (52) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GTGTAGATCT TCTGGTGCAC AGG

15 (53) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GCAATGCAGG TCATAGTGAG C

(54) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: YES

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:
 TGAGCATG TGACGGAAAT GCAGAG
 27
- (55) INFORMATION FOR SEQ ID NO:54:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: YES
 10
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:
 GTGATGACGA GTCTACTGAG CGCCAG
 27
- (56) INFORMATION FOR SEQ ID NO:55:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: NO
 20
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:
 GCATGCGAG CGCTTACAT TAC
 23
- (57) INFORMATION FOR SEQ ID NO:56:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: YES
 30
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:
 TTGGGTACA ATCTGAGGG CA
 22

- (58) INFORMATION FOR SEQ ID NO:57:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: NO
 5
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:
 ACTCCGTGTC CAGCAGACT CTG
 23
- (58) INFORMATION FOR SEQ ID NO:58:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 24 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: YES
 15
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:
 TCGGTGTTC TGACCCCTCA CGTG
 24
- (58) INFORMATION FOR SEQ ID NO:59:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
 (iv) ANTI-SENSE: NO
 25
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:
 CAGGCTTGG ATTATATGT CAGGATGG
 29
- (61) INFORMATION FOR SEQ ID NO:60:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs

-56-

- (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GGAGAGTCAG CTCGAAAGA ATTACGG

(62) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

TGATGTGATG CCGATACATA ATAGCAC

(63) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCTGATTCAT TTAGGTGAGA TTGAGAC

(64) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

-57-

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

CCCAAGCTTC CCCAGGTGTA TTGAT

(3) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GTGGATCCA CATAATGCAT TTCTC

(66) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

ATGATTCTCA ACTCTTCTAC TGAAGATGGT ATTTAAAGAA TCCAGATGA TTGTCCAAA 60

GCTGGAAGGC ATAATTACAT ATTGTTCATG ATTCTACTT TATACAGTAT CATCTTTGTG 120

GTGGGAATAT TTGGAACAG CTGTGTGTG ATAGTCATTT ACTTTTATAT GAAGCTGAAG 120

ACTGTGCCA GTGTTTCT TTTGAATTA GCACGTGGCTG ACTTATGCTT TTTACTGACT 240

25 TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACCGCT GSCCCTTTGG CAATTACCTA 300

TGTAAGATTG CTTACGCCAG CGTCAGTTTC AACCTGTACG CTAGTGTGTT TCTACTCAGC 360

TGTCTCAGCA TTGATCGATA CCTGGCTATT GTTCACCCAA TGAAGTCCCG CTTTCGAGGC 420

ACAATGCTTG TAGCCAAAGT CACCTGCATC ATCATTTGSC TGCTGGCAGG CTTGGCCAGT 480

TTGCCAGCTA TAATCCATCG AATGTATTT TTCAATGAGA ACACCAATAT TACAGTTTGT 540

30 GCTTTCCATT ATGAGTCCCA AAATTCAACC CTTCCGATAG GGCTGGGCCT GACCAAAAT 600

-58-

ATACAGGGT TCCTGTTCC TTTCGATC ATCTACAA GTTACTCT TATTGGAG 660
 GCCCTAAGA AGCTTATGA ATTCGAGG AACAAACCA GAATATGA TATTTTAA 720
 ATATATAG CATGAGCT TTCTTTTC TTTCCTGGA TTCCCAACA AATATGACT 780
 TTCTGATG TATGATCA ACTAGGATC ATAGGACT GTAGATTC AGATTTGTG 840
 5 GACAGGCCA TGCCATGAC CATTGATA GCTTATTTA ACAATTGCT GAATCTCTT 900
 TTTTATGCT TTCTGGGGA AAATTTAA AGATTTTC TCCAGCTCT AAATATAT 960
 CCCCCAAG CCAATCCCA CTCAACTT TCACAAAA TGAGCAGCT TTCTTACGC 1020
 CCTCAGATA ATGTAAGCT ATCCACCAAG AAGCTGCAC CATGTTTGA GTTGGATGA 1080

(67) INFORMATION FOR SEQ ID NO:66:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:66:

1 Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 15
 5
 20 Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 25
 30 Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 35
 40
 50 Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 55
 60
 25 Val Phe Leu Leu Asn Leu Leu Ala Asp Leu Cys Phe Leu Leu Thr 70
 75
 65 Leu Pro Leu Tyr Ala Val Tyr Thr Ala Met Glu Tyr Arg Tyr Pro Phe 85
 90
 100 Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Asn Leu 105
 110
 30 Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 115
 120
 Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val

-59-

130 135 140
 Ala Lys Val Thr Cys Ile Ile Ile Tyr Leu Leu Ala Gly Leu Ala Ser 160
 145 150 155
 Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 175
 165 170
 5 Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 185
 190
 Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe 205
 195 200

10 Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Tyr Lys Ala Leu Lys Lys 210
 215
 Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Asp Ile Phe Lys 230
 235
 15 Ile Ile Met Ala Ile Val Leu Phe Phe Phe Phe Ser Tyr Ile Pro His 245
 250
 Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg 265
 270
 Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr Ile 280
 275 285

20 Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly Phe 290
 295
 Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile 310
 315
 25 Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr 325
 330
 Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Ser Thr Lys Lys Pro 340
 345
 Ala Pro Cys Phe Glu Val Glu 355

(68) INFORMATION FOR SEQ ID NO:67:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

ACCATGGGCA GCCCTGGAA CGGCAGC

(69) INFORMATION FOR SEQ ID NO:68:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

AGAACACCA CCAGCAGGAC GCGGACGGTC TGCCGGTGG

(70) INFORMATION FOR SEQ ID NO:69:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

20 GTCCGCGTCC TGCTGGTGGT GGTCTGGCA TTTATAATT

(71) INFORMATION FOR SEQ ID NO:70:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CCTGGATCTT TATCCATCG TCTTCAGTT AGC

30 (72) INFORMATION FOR SEQ ID NO:71:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

5 CTGGAATTCT CCTGCCAGCA TGGTGA
 26

(73) INFORMATION FOR SEQ ID NO:72:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GCAGATCCT ATATTGCGTG CTCTGTCCCC
 30

(74) INFORMATION FOR SEQ ID NO:73:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 999 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

ATGGTGAACT CCACCACCG TGGATGAC ACTTCTCTGC ACTCTGAA CGCAGCAGT 60
 TACAGACTGC ACAGCAATGC CAGTGGTCC CTGGAAAG GCTACTCTGA TGGAGGGTGC 120
 TAGAGAGAAC TTTTGTCTC TCTGAGGTG TTTGTGACTC TGGGTGTCAT CAGCTTGTG 180
 GAGATACTCT TAGTGATTGT GCAATAGCC AGAACAAGA ATTCGATTC ACCCATGTAC 240
 30 TTTTTCATCT GCAGCTTGGC TGTGGCTGAT ATGCTGGTGA CGTTTCAA TGGATCAGAA 300
 ACCATTATCA TCACCTATT AAACAGTACA GATACGGATG CACAGAGTTT CACAGTGAAT 360
 ATTGATAATG TCATTGACTC GGTGATCTGT AGCTCTTGGC TTGCATCCAT TTGCAGCCTG 420

- 62 -

CTTCAATTG CAGTGCAGAG GACTTACT ACTCTTANG CTCCTCCAGTA CCATACATT 480
 ATGACAGTTA AGCGGGTGG GATCAGCATTA AGTTATCT GGGCAGCTTG CAGGTTTCA 540
 GGCAATTTGT TCATCATTTA CTCGATAGT AGTCGTGTCA TCATCTGCTT CATTACAGAG 600
 TCTTCACCA TGTGTGCTCT CAGGCTTCT CTCATGCTC ACATGTTCTT GATGCGCAGG 660
 5 CTCACATTA AGAGATATGC TGTCTCTCC GGCACGTGGT CCATCCGCCA AGGTGCCAAT 720
 ATGAGGGAG CGATTACTT GACCATCTG ATTTGAGCTCT TGTGTTCTG CTGGGCCCCA 780
 TCTTCTCTCC ACTTAATTT CTACATCTCT TGTCTCCAGA ATCCATATTG TGTGTGCTTC 840
 ATGCTCAGCT TTAATCTGTA TCTCATCTG ATCAGTGTGA ATTCAATCAT CGATCTCTG 900
 ATTATGAC TCCGAGTCA AGAAGTGGG AAAACCTTCA AAGATATCAT CTGTTGCTAT 960
 10 CCCCCTGGAG GCGTTGTGTA CTGTCTAGC AGATATTTA 999

(75) INFORMATION FOR SEQ ID NO:74:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 332 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:74:

Met Val Asn Ser Thr His Arg Gly Met His Thr Ser Leu His Leu Trp 15
 1 5
 Asn Arg Ser Ser Tyr Arg Leu His Ser Asn Ala Ser Glu Ser Leu Gly 30
 20
 Lys Gly Tyr Ser Asp Gly Gly Cys Tyr Glu Glu Leu Phe Val Ser Pro 45
 35
 Glu Val Phe Val Thr Leu Gly Val Ile Ser Leu Leu Glu Asn Ile Leu 60
 25 50
 Val Ile Val Ala Ile Ala Lys Asn Lys Asn Leu His Ser Pro Met Tyr 80
 65 70
 Phe Phe Ile Cys Ser Leu Ala Val Ala Asp Met Leu Val Ser Val Ser 95
 85
 Asn Gly Ser Glu Thr Ile Ile Ile Thr Leu Leu Asn Ser Thr Asp Thr 110
 100
 Asp Ala Glu Ser Phe Thr Val Asn Ile Asp Asn Val Ile Asp Ser Val 125

- 63 -

115 120 125
 Ile Cys Ser Ser Leu Leu Ala Ser Ile Cys Ser Ser Leu Ser Ile Ala 140
 130 135
 Val Asp Arg Tyr Phe Thr Ile Phe Tyr Ala Leu Glu Tyr His Asn Ile 155
 145 150
 Met Thr Val Lys Arg Val Gly Ile Ser Ile Ser Cys Ile Trp Ala Ala 175
 165
 Cys Thr Val Ser Gly Ile Leu Phe Ile Ile Tyr Ser Asp Ser Ser Ala 190
 180
 10 Val Ile Ile Cys Leu Ile Thr Met Phe Phe Thr Met Leu Ala Leu Met 205
 195
 Ala Ser Leu Tyr Val His Met Phe Leu Met Ala Arg Leu His Ile Lys 220
 210
 Arg Ile Ala Val Leu Pro Gly Thr Gly Ala Ile Arg Glu Gly Ala Asn 240
 225 230
 Met Lys Gly Ala Ile Thr Leu Thr Ile Leu Ile Gly Val Phe Val Val 255
 245
 Cys Trp Ala Pro Phe Phe Leu His Leu Ile Phe Tyr Ile Ser Cys Pro 270
 260
 15 Glu Asn Pro Tyr Cys Val Cys Phe Met Ser His Phe Asn Leu Tyr Leu 285
 275
 Ile Leu Ile Met Cys Asn Ser Ile Ile Asp Pro Leu Ile Tyr Ala Leu 300
 290
 Arg Ser Glu Glu Leu Arg Lys Thr Phe Lys Glu Ile Ile Cys Cys Tyr 320
 305 310
 Pro Leu Gly Gly Leu Cys Asp Leu Ser Ser Arg Tyr 330
 325

(76) INFORMATION FOR SEQ ID NO:75:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:75:

CCGAGCTTC GAGCTGAGTA AGCGGCGGG CT

(77) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTGGAATCA TTGCGCTGC CTCACCCCC A

10 (78) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1344 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

ATGAGCTGC TAAAGCTGAA CCGAGCGTG CAGGAAACG GACCCGGCC GGGGCTTCC 60
 CTGTGCGGC CGGGGGGCC TTCTCTCAAC AGCAGCAGTG TGGGCAACT CAGCTGCGAG 120
 20 CCCCCTCGCA TTGCGGAGC CGGACACGA GAATTGGAGC TGGCCATTAG AATCACTTT
 TAGCGAGTGA TCTTCTCTGAT GAGCGTTGGA GGAATATATG TCATCATCGT GGTCTGGGA 240
 CTGAGCGGCC GCTTGAGGAC TGTCAACCAT GCCTTCTCTC TTCTACTGGC AGTCAGGAC 300
 CTCCTGCTGG CTGTGCTTG CATGCCCTTC ACCCTCTTGC CCATCTCAT GGGCACATTC 360
 ATCTTTGGCA CCGTCATCTG CAAGCGGTT TCTTACTCTA TGGGGGTGTC TGTGAGTGTG 420
 25 TCCACGCTAA GCCTCTGTGC CATCGGCTGT GAGCGATATA GCGCCATCTG CCGACCACTG 480
 CAGGCACGAG TGTGCGAGAC GCGCTCCAC CCGGCTCGCG TGAATTGAGC CACGTGGCTG 540
 CTGTGCGGAC TACTCATNGT GCCTACACCC GTGTACACTG TCGTCAACCG AGTGGGCTT 600
 CGTGTGCTGC AGTGGGTGCA TCGCTGGGCC AGTGGCGGG TCGGCCAGAC CTGGTCCGTA 660
 CTGTGCTTTC TGCTCTTGT TTTCATCCCA GGTGTGGTTA TGGCGTGGC CTACGGGCTT 720
 30 ATCTCTCGCG AGCTCTACTT AGGGTTTCGC TTGACGCGG ACAGTGACAG CGACAGCCAA 780
 AGCAGGCTCC GAACCAAGG CCGGCTGCCA GGGGCTGTTT ACCAGAACCG CGGTGCGCG 840

CTGAGACTG GCGCGTTGG CAAAGACAGC GATGGCTGCT ACGTGCAACT TCCAGTTCC 900
 CGGCTGCCC TGGAGCTGAC GGGGCTGAG GCTCTGGGC CGGATCCGG CTCGCGGCC 960
 ACCAGGCCA AGCTGCTGGC TAAGAAGCG GTGGTGGAA TGTGCTGGT GATCGTTGTG 1020
 CTTTTTTC TGTGTTGTTT GCCAGTTAT AGTGCCAACA CGTGGCGCG CTTTGA/NGC 1080
 5 CCGGGTGAC ACCGAGCACT CTCGGGTGCT CCTATCTCT TCATTCACAT GCTGAGCTAC 1140
 GCTCGGCT GTGTCAACCC CCGGTCTAC TCGTTCATGC ACCGTGCTT TCGCAGGCC 1200
 TGCCTGAAA CTGCGCTCG CTGCTGCCCC CGGCTCCAC GAGCTGCCCC CAGGGCTCTT 1260
 CCCGATGAG ACCCTCCAC TCCCTCCAT GCTTCGCTGT CCAGGCTTAG CTACACCAC 1344
 ATCAGCACAC TGGGCCCTGG CTGA

10 (79) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 447 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

Met Glu Leu Leu Lys Leu Asn Arg Ser Val Gln Gly Thr Gly Pro Gly 15
 1
 Pro Gly Ala Ser Leu Cys Arg Pro Gly Ala Pro Leu Leu Asn Ser Ser 30
 20
 Ser Val Gly Asn Leu Ser Cys Glu Pro Pro Arg Ile Arg Gly Ala Gly 45
 35
 Thr Arg Glu Leu Glu Leu Ala Ile Arg Ile Thr Leu Tyr Ala Val Ile 60
 50
 Phe Leu Met Ser Val Gly Gly Asn Met Leu Ile Ile Val Val Leu Gly 80
 65
 Leu Ser Arg Arg Leu Leu Thr Val Thr Asn Ala Phe Leu Leu Ser Leu 95
 85
 Ala Val Ser Asp Leu Leu Leu Ala Val Ala Cys Met Pro Phe Thr Leu 110
 100
 Leu Pro Asn Leu Met Gly Thr Phe Ile Phe Gly Thr Val Ile Cys Lys 125
 115 120

- 66 -

Ala Val Ser Tyr Leu Met Gly Val Ser Val Ser Val Ser Thr Leu Ser
 130 135 140
 Leu Val Ala Ile Ala Leu Glu Arg Tyr Ser Ala Ile Cys Arg Pro Leu
 145 150 155 160
 Gln Ala Arg Val Trp Gln Thr Arg Ser His Ala Ala Arg Val Ile Val
 165 170 175
 Ala Thr Trp Leu Ser Gly Leu Leu Met Val Pro Tyr Pro Val Tyr
 180 185 190
 Thr Val Val Gln Pro Val Gly Pro Arg Val Leu Gln Cys Val His Arg
 195 200 205
 Trp Pro Ser Ala Arg Val Arg Gln Thr Trp Ser Val Leu Leu Leu
 210 215 220
 Leu Leu Phe Phe Ile Pro Gly Val Val Met Ala Val Ala Tyr Gly Leu
 225 230 235 240
 Ile Ser Arg Glu Leu Tyr Leu Gly Leu Arg Phe Asp Gly Asp Ser Asp
 245 250 255
 Ser Asp Ser Gln Ser Arg Val Arg Asn Gln Gly Gly Leu Pro Gly Ala
 260 265 270
 Val His Gln Asn Gly Arg Cys Arg Pro Glu Thr Gly Ala Val Gly Lys
 275 280 285
 Asp Ser Asp Gly Cys Tyr Val Gln Leu Pro Arg Ser Arg Pro Ala Leu
 290 295 300
 Glu Leu Thr Ala Leu Thr Ala Pro Gly Pro Gly Ser Gly Ser Arg Pro
 305 310 315 320
 Thr Gln Ala Lys Leu Leu Ala Lys Lys Arg Val Val Arg Met Leu Leu
 325 330 335
 Val Ile Val Val Leu Phe Phe Leu Cys Trp Leu Pro Val Tyr Ser Ala
 340 345 350
 Asn Thr Trp Arg Ala Phe Asp Gly Pro Gly Ala His Arg Ala Leu Ser
 355 360 365
 Val Ala Pro Ile Ser Phe Ile His Leu Leu Ser Tyr Ala Ser Ala Cys
 370 375 380
 Val Asn Pro Leu Val Tyr Cys Phe Met His Arg Arg Phe Arg Gln Ala
 385 390 395 400
 Cys Leu Glu Thr Cys Ala Arg Cys Cys Pro Arg Pro Pro Arg Ala Arg
 405 410 415
 Pro Arg Ala Leu Pro Asp Glu Asp Pro Pro Thr Pro Ser Ile Ala Ser

- 67 -

Leu Ser Arg Leu Ser Tyr Thr Thr Ile Ser Thr Leu Gly Pro Gly
 420 425 430 435 440 445
 (80) INFORMATION FOR SEQ ID NO:79:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO:79:
 TGCAGCTTA AAAAGAAA AATGACAC
 (81) INFORMATION FOR SEQ ID NO:80:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO:80:
 TAAAGATCCC TTCCTTCA AACATCCTTG
 (82) INFORMATION FOR SEQ ID NO:81:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1014 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO:81:
 ATGACACGCA CATGNTTGA AGACACGAT GACCTGGATC ACTATTTGTT TCCCATTTT 60
 TACATCTTTG TGATTTAGT CAGATTCGA GCCAATATTG GATCTCTGAG TGTGTCTTC 120
 CTGCACGCA AGAAGGAAA TGACTAGGA ATTACTCTT TCAATTTGTC ACTATCAGAT 180
 TTACTTAG CATTAAGT CCGTTTAGG ATTGATTATA CTGGATTAA AGACACTGG 240

5 ACTTCTCTC CTGCTCTGTG CAAAGGAGT GCCTTTCTCA TGTACATGAA GTTTTACAGC 300
 AGCAGAGCAT TCCTCACCTG CATTCGCCGT GATCGGTATT TGGCTGTGTG CTACCCCTTG 360
 AAGTTTTTTT TCCTAAGAC AAGAGAANT GCACATGCG TCAGCTGTGC CATCTGGATA 420
 TTGGAACCA TCTTCAATGC TGTCTATGTC TGGGAGATG AACAGTTGT TGAATATGTC 480
 5 GATGCCGAAA AGTCTAATTT TACTTTATGC TATGACAAAT ACCCTTTAGA GAAATGCCAA 540
 ATCAACCTCA ACTTGTTCAG GAGGTGTACA GGCTATGCAA TACCTTTTGGT CACCATCCTG 600
 ATCTGTAAAC GGAAGATCTA CCAAGCTGTG CGGCACAATA AAGCCACGGA AACACAGGAA 660
 AAGAAGAGAA TCTATAAACT ACTTGTACGC ATCAGAGTTA CTTTGTGCTT ATGCTTTACT 720
 CCCTTTCATG TGAATGTGCT GATTCGCTGC ATTTTAGAGC ATGCTGTGAA CTTCGAAGAC 780
 10 CACAGCAANT CTGGGAGCG AACCTACACA ATGTATAGAA TCACGGTTGC ATTACAGAT 840
 TTTAAATGTG TTGCTGATCC AATCTGTAC TGTTTTGTTA CCGAAACAGG AAGATATGAT 900
 ATGTGGAATA TATTAAATTT CTGCACTGGG AGGTGTAAATA CATCACAAAG ACAAGGAAA 960
 CGCATACTTT CTGTGTCTAC AAAGATACT ATGGAAATAG AGGTCCTTGA GTAG 1014

(83) INFORMATION FOR SEQ ID NO:82:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 337 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 20 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

1 Met Asn Ser Thr Cys Ile Glu Gln His Asp Leu Asp His Tyr Leu 15
 20 Phe Pro Ile Val Tyr Ile Phe Val Ile Ile Val Ser Ile Pro Ala Asn 30
 25 Ile Gly Ser Leu Cys Val Ser Phe Leu Gln Pro Lys Lys Glu Ser Glu 45
 50 Leu Gly Ile Tyr Leu Phe Ser Leu Ser Leu Ser Asp Leu Leu Tyr Ala 60
 65 Leu Thr Leu Pro Leu Trp Ile Asp Tyr Thr Trp Asn Lys Asp Asn Trp 75
 80 Thr Phe Ser Pro Ala Leu Cys Lys Gly Ser Ala Phe Leu Met Tyr Met

85 90 95
 Lys Phe Tyr Ser Ser Thr Ala Phe Leu Thr Cys Ile Ala Val Asp Arg 110
 100 105 110
 Tyr Leu Ala Val Val Tyr Pro Leu Lys Phe Phe Leu Arg Thr Arg 125
 115 120 125
 Arg Ile Ala Leu Met Val Ser Leu Ser Ile Trp Ile Leu Glu Thr Ile 140
 130 135 140
 Phe Asn Ala Val Met Leu Trp Glu Asp Glu Thr Val Val Glu Tyr Cys 155
 145 150 155 160
 Asp Ala Glu Lys Ser Asn Phe Thr Leu Cys Tyr Asp Lys Tyr Pro Leu 175
 165 170 175
 Glu Lys Trp Gln Ile Asn Leu Asn Leu Phe Arg Thr Cys Thr Gly Tyr 190
 180 185 190
 Ala Ile Pro Leu Val Thr Ile Leu Ile Cys Asn Arg Lys Val Tyr Gln 205
 195 200 205
 Ala Val Arg His Asn Lys Ala Thr Glu Asn Lys Glu Lys Lys Arg Ile 220
 210 215 220
 Ile Lys Leu Leu Val Ser Ile Thr Val Thr Phe Val Leu Cys Phe Thr 235
 225 230 235 240
 Pro Phe His Val Met Leu Leu Ile Arg Cys Ile Leu Glu His Ala Val 255
 245 250 255
 Asn Phe Glu Asp His Ser Asn Ser Gly Lys Arg Thr Tyr Thr Met Tyr 270
 260 265 270
 Arg Ile Thr Val Ala Leu Thr Ser Leu Asn Cys Val Ala Asp Pro Ile 285
 275 280 285
 Leu Tyr Cys Phe Val Thr Glu Thr Gly Arg Tyr Asp Met Trp Asn Ile 300
 290 295 300 305
 Leu Lys Phe Cys Thr Gly Arg Cys Asn Thr Ser Gln Arg Gln Arg Lys 315
 305 310 315 320
 Arg Ile Leu Ser Val Ser Thr Lys Asp Thr Met Glu Leu Glu Val Leu 335
 325 330 335
 Glu

(84) INFORMATION FOR SEQ ID NO:83:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 40 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:83:

5 CAGGAGAAG AACGAGCTG TCATTATGAT GGTACAGTG
40

(85) INFORMATION FOR SEQ ID NO:84:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:84:

15 CACTGTACC ATCATATGTA CAGCTGTT CTCTCTCTG
40

(86) INFORMATION FOR SEQ ID NO:85:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:85:

25 GGCACCGGC AGACCAACG CGTCTGCTG
30

(87) INFORMATION FOR SEQ ID NO:86:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:86:

CTCCTTGGT CCTCCTATG TTGTCAAG T
31

(88) INFORMATION FOR SEQ ID NO:87:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:87:

GGAAGAAG AGATCAAA AACTCTGT CAGCATC

(89) INFORMATION FOR SEQ ID NO:88:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:88:

20 CTCCTTGGT CCTCTATG TTGTCAAG T

(90) INFORMATION FOR SEQ ID NO:89:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:89:

ATGATTCTCA ACTCTTCTAC TGAGATGAT ATTAAGAAG TCCAGATGA TTGCCCAA 60
AGTATGAGCA GTGTTTCTCT TTGAATTTA GCACGAGCTG ACTTATAT GAAGCTGAG 120
GTGGGATAT TTGAAGAAG CTGAGTGTG ATGATCATTT ACTTTTAT GAAGCTGAG 180
ACTGTGACCA GTGTTTCTCT TTGAATTTA GCACGAGCTG ACTTATAT TTACTGACT 240
TTGCACATAT GGCTGTCTA CACAGCTATG GAATACCGCT GGCCCTTGG CAATTAACCTA 300

- 72 -

TGTAAAGATTG CTTACAGCCAG CGTCAGTTTC AACCTGTACG CTAGTGTGTT TCTACTCAGC 360
 TGTCTCAGCA TTGATCGATA CCGTGGCTATT GTTCACCCAA TGAAGTCCCG CTTTCGACGC 420
 ACAATGCTTG TAGCCAAAGT CACCTGCATC ATCATTGGCG TGCTGGCAGG CTTGGCCAGT 480
 TTGCCAGCTA TAATCCATCG AAATGTATTT TTCAATGAGA ACACCAATAT TACAGTTTGT 540
 5 GCCTTCCATT ATGAGTCCCA AAATCAACG CTTCCGATAG GCGTGGSCCT GACCAAAAAT 600
 ATACTGGGTT TCGTGTTC TTTTCTGATC ATCTTACAA GTTATACTCT TATTTGGAAG 660
 GCCCTAAGA AGGCTTATGA AATTCAAGG AACAAACCAA GAAATGATGA TATTAAGAAG 720
 ATAATTATGG CAATTGTGCT TTTCTTTTTC TTTTCTGGA TTCCCCACCA AATATTCACT 780
 TTTCTGGATG TATTGATTC ACTAGGCATC ATAGCTGACT GTAGAAATGC AGATATTGTG 840
 10 GACACGECGA TGCTATATCAC CATTTGTATA GCTTATTTTA ACAATTGCTT GAATCTCTTT 900
 TTTTATGGCT TTCTGGGGAA AAAATTAAAG AGATATTTTC TCCAGCTTCT AAAATATATT 960
 CCCCCAAAG CCAATCCCA CTCAACTT TCACGAAA TGAGCACGCT TTCTTACCGC 1020
 CCTCAGATA ATGTAAAGCT ATCCACCAAG AAGCTGCGAC CATGTTTGA GGTGAGTGA 1080

(91) INFORMATION FOR SEQ ID NO:90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 359 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 15
 1
 Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 30
 20
 Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 45
 35
 Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 60
 50
 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 80
 65 70 75
 Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe

- 73 -

Gly Asn Tyr Leu Cys Lys Ile Ala Ser Val Ser Phe Asn Leu 95
 100 105 110
 Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 125
 115 120
 Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Thr Met Leu Val 140
 130 135 140
 Ala Lys Val Thr Cys Ile Ile Ile Trp Leu Leu Ala Gly Leu Ala Ser 155 160
 145 150 155
 Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 170 175
 165 170
 Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 185 190
 180 185
 Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe 205
 195 200 205
 Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Trp Lys Ala Leu Lys Lys 220
 210 215 220
 Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Ile Lys Lys 240
 225 230 235 240
 Ile Ile Met Ala Ile Val Leu Phe Phe Phe Ser Trp Ile Pro His 255
 245 250 255
 Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg 270
 260 265 270
 Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr Ile 285
 275 280 285
 Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly Phe 300
 290 295 300
 Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile 320
 305 310 315 320
 Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr 335
 325 330 335
 Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Ser Thr Lys Lys Pro 350
 340 345 350
 Ala Pro Cys Phe Glu Val Glu 355

(92) INFORMATION FOR SEQ ID NO:91:

- 74 -

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

5 (11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:91:

CCAGAAATG ATGATATTA AAGATAAT ATGGC

(93) INFORMATION FOR SEQ ID NO:92:

- 10 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

15 (11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:92:

CTCCTCGGT CCTCCATCG TTGCAAG T

(94) INFORMATION FOR SEQ ID NO:93:

- 20 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1080 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO:93:

ATGATTTCA ACTTCTTAC TGAAGATGT ATTAAAGAA TCCAGATGA TTGTCCTAAA 60
 GCTGGAAGC ATATATACAT ATTTCATG ATTCTACT TATACAGAT CATCTTGTG 120
 GTGGAAAT TTGAAGACAG CTGGTGTG ATAGCAT TT ACTTTATAT GAAGCTGAAG 180
 ACTGTGCCA GTGTTTCT TTGATATTA GCACGTGCTG ACTTATGCT TTACTGACT 240
 TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACCGCT GGGCTTTG CATTACTCTA 300
 TGTAAATG CTTCAGCAG COTGAGTTT GCCCTGACG CTAGTGTG TCTACTACG 360
 TGTCTACGA TTGATGATA CCTGCTATT GTTCACCAA TGAATCCG CTTGAGGC 420

- 75 -

ACATGCTTG TAGCCAAAGT CACCTGCATC ATCATTTGC TGTGGCAGG CTTCGCAGT 480
 TTGCCAGCTA TATTCATG AATGATTT TTGATGAGA AACCCAAAT TACAGTTGT 540
 GCTTCATT ATGATGCCA AATTCACC CTTCGATG GGTGGGCT GACCAAAAT 600
 ATACGGGT TCTGTTTC TTTCGATC ATCTTACA GTTAACTCT TATTGGAAG 660
 5 GCCCTAAGA AGCTTATGA AATTCAGAG ACGAACCA GAATGATGA TATTTTAA 720
 ATATATAG CAATTGTCT TTCTTTTC TTTCCTGGA TTCCACCA AATATTCAT 780
 TTTCGATG TATGATTA ACTAGCATC ATACGTACT GTAGATTC AGATATTTG 840
 GACAGGCCA TGCCTATAC CATTTGATA GCTATTTTA ACATTCCTT GAATCCCTT 900
 TTTATGCT TTCTGGGAA AAAATTTAA AGATATTTT TCCAGCTCT AAAATTAAT 960
 10 CCCCAGAG CCAATCCA CTCAACTT TCAACAAAA TGACAGCT TTCTACCCG 1020
 CCTCAGATA ATGTAAGTC ATCCACGAG AAGCTGCAC CATGTTTGA GTTGAATGA 1080

(95) INFORMATION FOR SEQ ID NO:94:

- 15 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 359 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:94:

Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 15
 1 Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 20
 Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 25
 Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 30
 25 Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 35
 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 40
 50 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 45
 65 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 50
 70 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 55
 85 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 60
 90 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 65
 95 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 70
 Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Ala Leu 75

- 76 -

5 Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 110
115 120 125
Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val 140
130 135 140
Ala Lys Val Thr Cys Ile Ile Ile Trp Leu Leu Ala Gly Leu Ala Ser 150
145 155 160
Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 170
165 175
10 Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 180
185 190
Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe 200
195 205
15 Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Trp Lys Ala Leu Lys Lys 210
215 220
Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Asp Ile Phe Lys 230
225 235 240
Ile Ile Met Ala Ile Val Leu Phe Phe Phe Ser Trp Ile Pro His 245
250 255
20 Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg 260
265 270
Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr Ile 275
280 285
25 Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly Phe 290
295 300
Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile 305
310 315 320
Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr 325
330 335
30 Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Thr Lys Lys Pro 340
345 350
Ala Pro Cys Phe Glu Val Glu 355

(97) INFORMATION FOR SEQ ID NO:95:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid

- 77 -

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA (genomic)
(iv) ANTI-SENSE: NO
5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:
CCCAAGCTTC CCCAGGTGTA TTGTAT
(97) INFORMATION FOR SEQ ID NO:96:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
10 (ii) MOLECULE TYPE: DNA (genomic)
(iv) ANTI-SENSE: YES
15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:
CCTGCAGCG AACTGACTC TGCGTGAAG
(98) INFORMATION FOR SEQ ID NO:97:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
20 (ii) MOLECULE TYPE: DNA (genomic)
(iv) ANTI-SENSE: NO
25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:
CTGTACGCTA GTGTGTTTCT ACTCAGTGT CTCAGCAITG AT
(99) INFORMATION FOR SEQ ID NO:98:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
30 (ii) MOLECULE TYPE: DNA (genomic)

-78-

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:

GTTGATGCCA CATAATGAT TTCTC

26

(100) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1080 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:

ATGATTTCTCA ACTCTCTAC TGAAGATGT ATTAAAAGA TCCAGATGA TTCTCCAAA 60
GCTGGAGGC ATAATTACAT ATTGTGATG ATTCCTACT TATACATAT CATCTTGTG 120
GTGGGAATAT TTGGAACAG CTGTGTGTG ATAGTCATT ACTTTATAT GAAGCTGAAG 180
ACTGTGGCCA GTGTTTCTT TTGAAATTA GCAGTGTG ACTTATGCTT TTACTGACT 240
TTGCCACTAT GGGCTGTCTA CACAGCTATG GAATACGGCT GGCCCTTGTG CAATTACTTA 300
TGTAAAGATG CTTCAGCCAG CGTCAGTTT AACCTTACG CTAGTGTGT TCTACTACAG 360
TGTCTACACA TTGATGATA CTTGCTATT GTTCACCCCA TGAAGTCCG CCTTCAGCG 420
ACAAATGCTG TAGCCAAAGT CACTGTATC ATCATTTGGC TGTGTGCAAG CTGGCCAGT 480
TTGCCAGCTA TAATCATG AAATGATTT TTCAATGAGA ACACCAATAT TACAGTTGT 540
GCTTTCATT ATAGTCCCA AAATCAACC CTTCGATG GGTGGGGCT GACCAAAAT 600
ATACTGGGT TCTGTCTCC TTTTCTGATC ATCTTACAA GTTATTTTGG AATTGAAAA 660
CACTACTGA AGAGGAATAG CTATGGGAG AACAGATTA CCGGTGACCA AGTTAAGAG 720
ATAATTATG CAATGTGCT TTCTTTTTC TTTTCTGGA TTCCCAACA AATATGACT 780
25 TTCTGATG TATGATTA ACTAGCATC ATAGTACT GTAGATTC AGATATG 840
GACAGGCCA TGCCTATAC CATTTGATA GCTTATTTA ACAATGCT GAATCTCTT 900
TTTATGTCT TTCTGGGGA AAATTTAA AGATATTTT TCCAGTCT AAATATAT 960
CCCCAAG CCMAATCCA CTCAACTT TCAACAAAA TGAAGAGCT TTCCATCCG 1020
CCCTGATA ATGTAGCTC ATCCACAG AGCTGAC CATGTTTA GTTAGATA 1080

-79-

(101) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:

Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 15
1 Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 5
Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro 20
Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 35
Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 50
Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Cys Phe Leu Leu Thr 65
Leu Pro Leu Tyr Ala Val Tyr Thr Ala Met Glu Tyr Arg Tyr Pro Phe 85
Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Asn Leu 100
Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 115
Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val 130
Ala Lys Val Thr Cys Ile Ile Ile Thr Leu Leu Ala Gly Leu Ala Ser 145
Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 165
Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 180
Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro 195
Leu Ile Ile Leu Thr Ser Tyr Phe Gly Ile Arg Lys His Leu Leu Lys 210

- 80 -

Thr Asn Ser Tyr Gly Lys Asn Arg Ile Thr Arg Asp Gln Val Lys Lys
225 230 235 240
Ile Ile Met Ala Ile Val Leu Phe Phe Phe Ser Trp Ile Pro His
245 250 255
Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile Arg
260 265 270
Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr Ile
275 280 285
Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly Phe
290 295 300
Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr Ile
305 310 315 320
Pro Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser Thr
325 330 335
Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Thr Lys Lys Pro
340 345 350
Ala Pro Cys Phe Glu Val Glu
355

(102) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:

TCCGAAATCC AAATACTT GTAGAATGA TCAGAAA

(103) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

- 81 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

AGATCTTAAG AAGATAATTA TGGCAATTGT GCT

(104) INFORMATION FOR SEQ ID NO:103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

AATTCGAAA CACTTACTGA AGACGATAG CTATGGGAAG AACAGGATAA CCCGTGACCA

AG

(105) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 62 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

TTAACTTGGT CACGGGTAT CCTGTCTTC CCATAGCTAT TCGTCTTCAG TAAGTGTTTT

CG

(106) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1083 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

ATGATCTCA ACTCTCTAC TGAAGATGT ATTAAGA TCCAAAGTA TGTCCCAA 60
 GCTGGAAGC AATATACAT ATTGTGATG ATTCTACTT TATACAGTAT CATCTTGTG 120
 GTGGGAATAT TTGAAACAG CTGTGTGTG ATATCATTT ACTTTATAT GAAGCTGAG 180
 ACTGTGGCA GTGTCTTCT TTGAAATTA GCACTGGCTG ACTTANGCT TTACTGACT 240
 TGGCACTAT GGGCTGTCTA CACAGCTATG GAATACGCT GGGCTTTGG CATTAAGCTA 300
 TGTAAATTT CTTCAGCCAG CGTCACTTC AACCTGTAG CTATGTGTT TCTACTCAG 360
 TGTCTAGCA TTGATGATA CTTGGCTATT GTTCACCCA TGAAGTCCG CCTTGAGCG 420
 ACAATCTTG TAGCCAAAT CACTGTATC ATCATTTGG TGTGTGAGG CTGTGGCAGT 480
 TTGCGAGCTA TAATCCATG AATATATTT TTCAATGGA ACACCAATAT TACAGTTGT 540
 10 GCTTTCATTT ATGATGCCA AATTCACCC CTTCGATAG GGTGGGCTT GACCAAAAT 600
 ATATCTGGTT TCTGTCTCC TTCTGTATC ATCTTACCA GTTATACCTT TATTTGAGG 660
 GCCCTAAGA AGCTTATGA AATTCAGAG AACCAACCA GAATGATGA TATTTTAA 720
 AATATATG CAGCATGTT GCTTCTTCT TTCTTCTCT GATTTCCCA CCAATATTC 780
 ACTTCTG ATGTATGAT TCAACTAGC ATCATAGTG ACTGTAGAT TCGATATTT 840
 15 GTGACAGG CAGTCTAT CACCATTTGT ATAGCTTAT TTAACATTT CCTGAATCT 900
 CTTTATATG GCTTCTGGG GAAAAATTT AAAAGATTT TTCTCAGCT TCTAAATAT 960
 ATTCCTCCA AAGCAATC CCACTGAAAC CTTCACCAA AATGAGCAC GCTTCTTAC 1020
 CGCCCTCAG AATATGATG CTCACTCAC AAGAGCTG CACCATTTT TGAAGTGA 1080
 TGA 1083

20 (107) INFORMATION FOR SEQ ID NO:106:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 360 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULAR TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:106:

1 Met Ile Leu Asn Ser Ser Thr Glu Asp Gly Ile Lys Arg Ile Gln Asp 15
 5
 30 Asp Cys Pro Lys Ala Gly Arg His Asn Tyr Ile Phe Val Met Ile Pro

20 Thr Leu Tyr Ser Ile Ile Phe Val Val Gly Ile Phe Gly Asn Ser Leu 30
 35
 5 Val Val Ile Val Ile Tyr Phe Tyr Met Lys Leu Lys Thr Val Ala Ser 40
 50
 10 Val Phe Leu Leu Asn Leu Ala Leu Ala Asp Leu Cys Phe Leu Leu Thr 55
 65
 15 Leu Pro Leu Trp Ala Val Tyr Thr Ala Met Glu Tyr Arg Trp Pro Phe 60
 85
 10 Gly Asn Tyr Leu Cys Lys Ile Ala Ser Ala Ser Val Ser Phe Asn Leu 90
 100
 15 Tyr Ala Ser Val Phe Leu Leu Thr Cys Leu Ser Ile Asp Arg Tyr Leu 105
 115
 20 Ala Ile Val His Pro Met Lys Ser Arg Leu Arg Arg Thr Met Leu Val 120
 130
 25 Ala Lys Val Thr Cys Ile Ile Ile Tyr Leu Leu Ala Gly Leu Ala Ser 135
 145
 30 Leu Pro Ala Ile Ile His Arg Asn Val Phe Phe Ile Glu Asn Thr Asn 140
 165
 35 Ile Thr Val Cys Ala Phe His Tyr Glu Ser Gln Asn Ser Thr Leu Pro 160
 180
 40 Ile Gly Leu Gly Leu Thr Lys Asn Ile Leu Gly Phe Leu Phe Pro Phe 175
 195
 45 Leu Ile Ile Leu Thr Ser Tyr Thr Leu Ile Trp Lys Ala Leu Lys Lys 200
 215
 50 Ala Tyr Glu Ile Gln Lys Asn Lys Pro Arg Asn Asp Asp Ile Phe Lys 210
 230
 55 Ile Ile Met Ala Ala Ile Val Leu Phe Phe Phe Ser Trp Ile Pro 235
 245
 60 His Gln Ile Phe Thr Phe Leu Asp Val Leu Ile Gln Leu Gly Ile Ile 240
 260
 65 Arg Asp Cys Arg Ile Ala Asp Ile Val Asp Thr Ala Met Pro Ile Thr 265
 275
 70 Ile Cys Ile Ala Tyr Phe Asn Asn Cys Leu Asn Pro Leu Phe Tyr Gly 280
 295
 75 Phe Leu Gly Lys Lys Phe Lys Arg Tyr Phe Leu Gln Leu Leu Lys Tyr 300
 310
 80 320

-84-

Ile Pro Lys Ala Lys Ser His Ser Asn Leu Ser Thr Lys Met Ser
 330 335
 Thr Leu Ser Tyr Arg Pro Ser Asp Asn Val Ser Ser Thr Lys Lys
 340 345 350
 Pro Ala Pro Cys Phe Glu Val Glu
 355 360

(108) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

CCCAAGCTTC CCCAGGTGTA TTGAT

(109) INFORMATION FOR SEQ ID NO:108:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

AAGACAATT GCTGCATAAT TATCTTAAAA ATATCATC

(110) INFORMATION FOR SEQ ID NO:109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

-85-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

AAGATAATTA TGGCAGCAAT TGTCCTTTTC TTTTCTTT

(111) INFORMATION FOR SEQ ID NO:110:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

GTTGGATCCA CATAATGCAT TTTCCTC

(112) INFORMATION FOR SEQ ID NO:111:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1344 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

ATGGAGCTGC TAAAGCTGAA CCGGAGCGTG CAGGGAACCG GACCCGGGCC GGGGGCTTCC 60
 CTGTGCCGCC CGGGGGCGCC TCTCTCAAC AGCAGCAGTG TGGGCAACCT CAGCTGCAG 120
 CCCCCTGCA TTGCGGAGC CGGACACGA GAATTGGAGC TGGCCATTAG AATCACTCTT 180
 TAGCAGTGA TTTTCCTGAT GAGCGTTGGA GGAATATATC TCATCATCGT GGTCTGGGA 240
 CTGAGCCGCC GCTTGAGGAC TGTCACCAT GCCTTCTCC TCTCACTGGC AGTCAGGAC 300
 CTCCTGCTGG CTGTGGCTTG CATGCCCTTC ACCCTCTGC CCAATCTCAT GGGCACATTC 360
 ATCTTTGGCA CCGTCATCTG CAAGCGGTTT TCTTACCTCA TGGGGGTGTC TGTGAGTGTG 420
 TCCAGCGTAA GCCTCGTGGC CATCGCACTG GAGGATATA GCGCCATCTG CCGACCACTG 480
 CAGGACAGAG TGTGGCAGAC GCGCTCCAC GCGCTCGCG TGATTGTAGC CACGTGGCTG 540
 CTGTCCGGAC TACTCATGCT GCCCTACCCC GTGTACACTG TCGTGAACCC AGTGGGGCCT 600
 CGTGTGCTGC AGTGCGTGCA TCGCTGSCCC AGTGGCGGG TCCGCCAGAC CTGGTCCGTA 660

CTGCTGCTTC TGCTCTTGT CTTCATCCCA GGTGTGTGTTA TGCCCTGTGC CTACGGGCTT 720
 ATCTCTGCG AGCTTACTT AGGCTTGGC TTGTAGCGCG ACGTGTACAG CGACAGCCAA 780
 AGCAGGGTCC GAACACAGG CGGCTGTCCA GGGGCTGTTC ACCAAGACGG GCGTTCGGG 840
 CCTGAGACTG GCGCGHTVGG CAAGAGACAG GATGTGTCTT ACATGCACTT TCCAGCTTCC 900
 CGGCTGTGCC TGAGCTGTAC GGGCTGTAGG GCTCTGTGGC CGGATCCGG CTCGGGCCC 960
 ACCAGAGCCA AGCTGTGTGC TTAGAGCGC GTAAAGCAA TGTGTGTGT GTATGTGTG 1020
 CTTTCTTTC TGTTGTGTT GCCAGTTTAT AGTCCACAA CGTGGCGCGC CTTTGTATGGC 1080
 CCAGGTGCAC ACCGAGCACT CTCGGGTGCT CCTATCTCT TCAATCACTT GCTGAGCTAC 1140
 GCGTGGGCTT GTGTCAACCC CTTGTCTAC TGCTTATGC ACCGTGCTT TCGCAGAGCC 1200
 TGCCGTGAAA CTTGGCTTGC CTGTGTGCG CGGCTGTCCAC GAGCTGTGCC CAGGGCTCTT 1260
 CCGATGTAGG ACCCTCCAC TCCCTCCANT GCTTGGCTGT CAGGCTTNG CTACAGCAC 1320
 ATCAGCACAC TGCGCTGTGC CTGA 1344

(113) INFORMATION FOR SEQ ID NO:112:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 447 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:112:

Met Glu Leu Leu Lys Leu Asn Arg Ser Val Gln Gly Thr Gly Pro Gly 1
 5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995

Ala Val Ser Asp Leu Leu Leu Ala Val Ala Cys Met Pro Phe Thr Leu 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995

- 88 -

385 390 395 400
 Cys Leu Glu Thr Cys Ala Arg Cys Cys Pro Arg Pro Arg Pro Arg Ala Arg
 405 410 415
 Pro Arg Ala Leu Pro Asp Glu Asp Pro Thr Pro Thr Pro Ser Ile Ala Ser
 420 425 430
 Leu Ser Arg Leu Ser Tyr Thr Thr Ile Ser Thr Leu Gly Pro Gly
 435 440 445

(114) INFORMATION FOR SEQ ID NO:113:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

CAGGAGCATG CGCTTCACGC GCTTCTTAGC CCAG

(115) INFORMATION FOR SEQ ID NO:114:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

25 AAGAGCCGCT GAAGCGCATG CTGCTGTGTA TCGTT

(116) INFORMATION FOR SEQ ID NO:115:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

- 89 -

ATGAGAGAAA GAATCAAAAAG AATGTTCTAT ATA
 (117) INFORMATION FOR SEQ ID NO:116:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 33 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

TATATAGAAC ATTCTTTTGA TTCCTTCTC CAT

(118) INFORMATION FOR SEQ ID NO:117:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: NO

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CGCTCTCTGG CCTTGAAGCG CACGCTCAGC

(119) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

GCTGAGCGTG CGCTTCAGG CCAGAGACCG

(120) INFORMATION FOR SEQ ID NO:119:

- (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:119:
 CCCAGGAAAA AGGTGAAAGT CAAAGTTTC

30

- 10 (121) INFORMATION FOR SEQ ID NO:120:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:120:
 GAAACTTGG ACTTCACCT TTTTCTGGG

30

- 20 (122) INFORMATION FOR SEQ ID NO:121:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:121:
 GGGGGCGGG TGAAACGGCT GGTGAGC

27

- 30 (123) INFORMATION FOR SEQ ID NO:122:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

- (11) MOLECULE TYPE: DNA (genomic)
 (1v) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:122:
 GCTCACGAC CGTTTCACCC GGGCCCC

27

- 5 (124) INFORMATION FOR SEQ ID NO:123:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: NO

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:123:
 CCCCTTGAAA AGCCTAAGAA CTTGTCATC

30

- 15 (125) INFORMATION FOR SEQ ID NO:124:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(1v) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:124:
 GATGACCAAG TTCTTAGGCT TTCAAGGGG

30

- 25 (126) INFORMATION FOR SEQ ID NO:125:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

30

- 92 -

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

GATCTCTAGA ATGAACAGCA CATGATTGA AG

(127) INFORMATION FOR SEQ ID NO:126:

5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(iv) ANTI-SENSE: YES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

CTAGGTACC CGCTCAAGGA CTTCTAATTC CATAG

(128) INFORMATION FOR SEQ ID NO:127:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1296 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

ATGACGGCC TTAACATTAC CCGGAGCAG TTCTCTCGGC TGTCTGGGA CCACACCTG 60

ACGCGGAGC AGTTCATCG TCCTGACCG CTGACCGC TCGTCTACAC CCACAGCTG 120

CCGGACGCG CCAAGCTGC CCTCGTGCTC ACCGGGTGC TCATCTTGC CCTGGCGTC 180

25 TTGGCAATG CTCTGTGTT CTAGTGTGT ACCCGCAGCA AGGCCATGCG CACCGTCACC 240

AACATCTTA TCTGCTCCTT GCGCTCAGT GACCTGCTCA TCACCTTCTT CTGCATTCCC 300

GTCAACATGC TCCAGAACAT TTCCGACAA TCAGTGGGG GTGCTTTTCAT TTGCAAGATG 360

GTGCCATTG TCCAGTCTAC CGCTGTGTG ACAGAAATGC TCACATATGAC CTGCATTGCT 420

GTGGAAGGC ACCAGGACT TGTGATCCT TTAAATGA AGTGGCAATA CACCAACCGA 480

- 93 -

AGGCTTTCA CAATCTAGG TGTGCTGG CTGTCGAG TCATCTAGG ATCACCCTG 540

TGCACGTGC AACAACTGA GATCAATAT GACTTCTAT ATGAAAAGGA ACACATCTGC 600

TGCTTAGAG AGTGACAG CCGTGTGC CAGAAGAT ACACACCTT CATCTTTGTC 660

ATCTCTTCC TCTGCTCT TATGTGATG CTTATCTGT ACAGTAAT TGGTTATGAA 720

5 CTTGGATAA AGAAAAGAT TGGGATGTT TCACTGCTT GAATATTCA TGGAAAAGAA 780

ATGTCMAAA TAGCCAGAA GAAGAAAGCA GCTAAGATTA TGATGTGAC AGTGTGGCT 840

CTCTTGTG TGTGTGGC ACCATTCCAT GTGTCCATA TGATGATTC ATACAGTAAT 900

TTTGAAGAG AATATGATA TGTCAATC AAGATGATTT TTGCTATCT GCAATTTAT 1020

GGATTTTCCA ACTCCATCTG TAATCCCAT GTCTGTGAT TTATGATGA AAACCTTCAA 1080

10 AAAATGTTT TGTCTGCACT TTGTTATGC ATAGTAAATA AAACCTTCTC TCCAGCACAA 1140

AGGCATGAA ATTCAGGAT TACATGATG CCGAAGAAAG CAAAGTTTTC CCTCAGAGAG 1200

AATCCAGTG AGGAACCCA AGGAGAAGCA TTCAGTGATG GCACATTTGA AGTCAATTTG 1260

TGTGAACAGA CAGAGAGAA GAAAAGCTC AAACGACATC TTGCTCTCT TAGGCTGTGA 1296

CTGCTGAGA ATTCTCTTT AGACAGTGGG CATTAA

15 (129) INFORMATION FOR SEQ ID NO:128:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 431 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS:

(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:128:

Met Gln Ala Leu Asn Ile Thr Pro Gln Phe Ser Arg Leu Leu Arg 15
1 5 1025 Asp His Asn Leu Thr Arg Gln Phe Ile Ala Leu Tyr Arg Leu Arg 30
20 25 30Pro Leu Val Tyr Thr Pro Gln Leu Pro Gly Arg Ala Lys Leu Ala Leu 45
35 40 4530 Val Leu Thr Gly Val Leu Ile Phe Ala Leu Ala Leu Phe Gly Asn Ala 55
50 55 60Leu Val Phe Tyr Val Thr Arg Ser Lys Ala Met Arg Thr Val Thr 80
65 70 75 80

-94-

Asn Ile Phe Ile Cys Ser Leu Ala Leu Ser Asp Leu Leu Ile Thr Phe
 85 90 95
 Phe Cys Ile Pro Val Thr Met Leu Gln Asn Ile Ser Asp Asn Trp Leu
 100 105 110
 Gly Gly Ala Phe Ile Cys Lys Met Val Pro Phe Val Gln Ser Thr Ala
 115 120 125
 Val Val Thr Glu Met Leu Thr Met Thr Cys Ile Ala Val Glu Arg His
 130 135 140
 Gln Gly Leu Val His Pro Phe Lys Met Lys Trp Gln Tyr Thr Asn Arg
 145 150 155 160
 Arg Ala Phe Thr Met Leu Gly Val Val Trp Leu Val Ala Val Ile Val
 165 170 175
 Gly Ser Pro Met Trp His Val Gln Gln Leu Glu Ile Lys Tyr Asp Phe
 180 185 190
 Leu Tyr Glu Lys Glu His Ile Cys Cys Leu Glu Glu Trp Thr Ser Pro
 195 200 205
 Val His Gln Lys Ile Tyr Thr Thr Phe Ile Leu Val Ile Leu Phe Leu
 210 215 220
 Leu Pro Leu Met Val Met Leu Ile Leu Tyr Ser Lys Ile Gly Tyr Glu
 225 230 235 240
 Leu Trp Ile Lys Lys Arg Val Gly Asp Gly Ser Val Leu Arg Thr Ile
 245 250 255
 His Gly Lys Glu Met Ser Lys Ile Ala Arg Lys Lys Lys Arg Ala Lys
 260 265 270
 Ile Met Met Val Thr Val Val Ala Leu Phe Ala Val Cys Trp Ala Pro
 275 280 285
 Phe His Val Val His Met Met Ile Glu Tyr Ser Asn Phe Glu Lys Glu
 290 295 300
 Tyr Asp Asp Val Thr Ile Lys Met Ile Phe Ala Ile Val Gln Ile Ile
 305 310 315 320
 Gly Phe Ser Asn Ser Ile Cys Asn Pro Ile Val Tyr Ala Phe Met Asn
 325 330 335
 Glu Asn Phe Lys Lys Asn Val Leu Ser Ala Val Cys Tyr Cys Ile Val
 340 345 350
 Asn Lys Thr Phe Ser Pro Ala Gln His Gly Asn Ser Gly Ile Thr
 355 360 365
 Met Met Arg Lys Lys Ala Lys Phe Ser Leu Arg Glu Asn Pro Val Glu

-95-

370 375 380
 Glu Thr Lys Gly Glu Ala Phe Ser Asp Gly Asn Ile Glu Val Lys Leu
 385 390 395 400
 Cys Glu Gln Thr Glu Glu Lys Lys Lys Leu Lys Arg His Leu Ala Leu
 405 410 415 420
 Phe Arg Ser Glu Leu Ala Glu Asn Ser Pro Leu Asp Ser Gly His
 425 430 435

(130) INFORMATION FOR SEQ ID NO:129:
 (1) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2040 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (11) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:
 ATGGGAGACC CCTGGAGACGG GAGGAGACGG CCCGAGGGGG GGGGGAGGCC GCGGTGGCCC
 60
 GCGCTGCGGC CTTCGAGACGA GCGCGCGTGC TCGCCCTTTC CCTGGGGGGC GCTGTGCGG
 120
 GTGACCGGCTG TGTGCTGTG CCTGTTCGTC GTGCGGGTGA GCGGACAGT GGTGACCGT
 180
 ATGCTGATCG GAGCTACCG GAGACATGCG ACCACACCA ACTTGTAAGT GGGCAGCATG
 240
 GCGGATCGG AACTTACTAT CTGTCTCGGG CTGCGATTGG ACCGTACCG CCTTGGCGG
 300
 TCGCGGACCT GGGTGTTCGG GCGGCTGTTC TCGCGCTGT CCTCTTACGT GGGCGAGGCG
 360
 TGGACCTTACG CGAGCTGTCT GCACATGACC GCGCTTACGG TCGAGGCTTA CTGGGCAATC
 420
 TGGCGCCCGC TCGGAGCCCG CGTCTTGGTC ACCGAGGACC GGTTCGCGGC GCTCATGCT
 480
 GTGCTCTGGG CGGTGAGCT GCTCTGTGCC GGTTCCTTCT TGTTCCTTGT GGGGTGAG
 540
 CAGAGACCCG GCAATTCGCT AGTCCGGAGC CTCATATGCA CCGAGGAGAT CGGCTCTCG
 600
 CCTCTGCTT CGTGGCGGCC TCTGTGATTC TCGGAGGAGC CACCGGCTTC CCGCGCGTGC

660
 720
 5
 780
 840
 900
 960
 1020
 1080
 1140
 1200
 1260
 1320
 1380
 1440
 1500
 1560
 1620
 1680
 1740

GGGCCGAGG CCGCGAGGC CGCGGCGTG TTACGCCCG AATGCCGCC GAGCCCGCG
 CAGCTGGCG CGCTGGGTGT CATGCTGCG GTACACCGG CCTACTTCTT CTTGCCCTTT
 CTGTGCTCA GCATCTCTTA CGGGCTCATC GGGCGGAGC TGTGAGCAG CCGCGCGCG
 CTGCGAGCC CGCGCGCTC GGGCGGCGG AGNGGCCACC GGCAGACCA ACGGCTCTG
 CTTAAGTGA GCCGCCGTG TTCCAAAGAC GCTGCTCTG AGTCCGCCCC GCCCGGAGC
 GCGCAAGGC TGGTCCCTT TCCCTGCTC GCCAGCTCT GGGGCGCGT TCCAGCTCC
 TTCTGATTT CGATTCCAGC CTCCACCCCG CGGTACTTCC CATCCCCCGA GAAACACATG
 TCCGTGCCCC CAGGAGCTCT GGGGAGCCCC AGGCGCTTT GAGGGTGGG TCCCGCGATC
 CGATTCACTA ACCAGCAGT CTTTCCAGA GCCTCTGGA CCAGAAAGGA GAGTTGTAA
 TTCTAATCC AACCACTGT TAGATCCAC AATGAGGAG TCTCACAGT GCTCTTGAGA
 AGACGAGGA GATTTCAITTA AGCTAATAIT TTTTATTTAA TGTTAAGTGA TGCTGAGGC
 TAAAGTAAAC CTTGCTCGTA TCAGAAAGTA AAGATTGTGC AGACTGTGT TAGAATTTCT
 TTCACAGAG AACAGAAAAC TTGTCTCGA AGTGGGTTG TGAAGGAAG CCGCCAGAG
 CGGCTGTTC AGAGAAATTT CTCTTCTGG TTTATGTCCA GCCTTGATTA CACATATGG
 AGCCTACTAT CGAGTTTAA AGCAAGTATC CATGACCTI GCAGCTGTGT CATTTTTCT
 GGGGTGAGGA TCTGCTAGG TAGAATTTT CTCTAATTTA TTTTGTGTT ACTTGTAAT
 GCAGATGTT CTTGTGCGG GTGGGGGTT TATTGTCTTC CCAATGCTTT TGTATATCC
 GGTGCTGTGT CTTATGTTC AGTGTGTTG GTTCTGCGAT TTATATTTG CTGGTGGCC

1800
 1860
 1920
 1980
 2040
 (131) INFORMATION FOR SEQ ID NO:130:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 412 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:
 Met Gly Ser Pro Trp Asn Gly Ser Asp Gly Pro Glu Gly Ala Arg Glu
 1 5 10 15
 Pro Pro Trp Pro Ala Leu Pro Pro Cys Asp Glu Arg Arg Cys Ser Pro
 20 25 30
 Phe Pro Leu Gly Ala Leu Val Pro Val Thr Ala Val Cys Leu Cys Leu
 35 40 45
 Phe Val Val Gly Val Ser Gly Asn Val Val Thr Val Met Leu Ile Gly
 50 55 60
 Arg Tyr Arg Asp Met Arg Thr Thr Thr Asn Leu Tyr Leu Gly Ser Met
 65 70 75 80
 Ala Val Ser Asp Leu Leu Ile Leu Leu Gly Leu Pro Phe Asp Leu Tyr
 85 90 95
 Arg Leu Trp Arg Ser Arg Pro Trp Val Phe Gly Pro Leu Leu Cys Arg
 100 105 110
 Leu Ser Leu Tyr Val Gly Glu Gly Cys Thr Tyr Ala Thr Leu Leu His
 115 120 125
 Met Thr Ala Leu Ser Val Glu Arg Tyr Leu Ala Ile Cys Arg Pro Leu
 130 135 140

- 98 -

Arg Ala Arg Val Leu Val Thr Arg Arg Arg Val Arg Ala Leu Ile Ala
145 150 155 160
Val Leu Trp Ala Val Ala Leu Leu Ser Ala Gly Pro Phe Leu Phe Leu
165 170 175
Val Gly Val Glu Gln Asp Pro Gly Ile Ser Val Val Pro Gly Leu Asn
180 185 190
Gly Thr Ala Arg Ile Ala Ser Ser Pro Leu Ala Ser Ser Pro Pro Leu
195 200 205
Trp Leu Ser Arg Ala Pro Pro Ser Pro Ser Gly Pro Gly Thr
210 215 220
Ala Glu Ala Ala Leu Phe Ser Arg Glu Cys Arg Pro Ser Pro Ala
225 230 235 240
Gln Leu Gly Ala Leu Arg Val Met Leu Trp Val Thr Thr Ala Tyr Phe
245 250 255
Phe Leu Pro Phe Leu Cys Leu Ser Ile Leu Tyr Gly Leu Ile Gly Arg
260 265 270
Glu Leu Trp Ser Ser Arg Arg Pro Leu Arg Gly Pro Ala Ala Ser Gly
275 280 285
Arg Glu Arg Gly His Arg Gln Thr Lys Arg Val Leu Leu Val Val
290 295 300
Leu Ala Phe Ile Ile Cys Trp Leu Pro Phe His Val Gly Arg Ile Ile
305 310 315 320
Tyr Ile Asn Thr Glu Asp Ser Arg Met Met Tyr Phe Ser Gln Tyr Phe
325 330 335
Asn Ile Val Ala Leu Gln Leu Phe Tyr Leu Ser Ala Ser Ile Asn Pro
340 345 350
Ile Leu Tyr Asn Leu Ile Ser Lys Lys Tyr Arg Ala Ala Phe Lys
355 360 365
Leu Leu Leu Ala Arg Lys Ser Arg Pro Arg Gly Phe His Arg Ser Arg
370 375 380
Asp Thr Ala Gly Glu Val Ala Gly Asp Thr Gly Gly Asp Thr Val Gly
385 390 395 400
Tyr Thr Glu Thr Ser Ala Asn Val Lys Thr Met Gly
405 410

35 (132) INFORMATION FOR SEQ ID NO:131:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1344 base pairs

- 99 -

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(1) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

ATGAGCTGC TAAAGTGAA CCGAGCGTG CAGGGAACG GACCCGGGCC GGGGGCTTCC
60
CTGTCCGCC CCGGGGGGCC TCTCTCAAC AGCAGCAGTG TGGGCAACT CAGCTGCAG
120
CCCCCTGCA TTGCGGAGC CGGACACGA GAATTGAGC TGGCCATTAG AATCACTCTT
180
TAGGAGTGA TCTTCTGAT GAGCGTTGA GGAATATAG TCATCATGT GGTCTGAGA
240
CTGAGCGGC GCGTGAAGC TGTACCAAT GCCTTCTCC TGTACTGGC AGTCAGGAC
300
CTCTGCTGG CTGTGGCTTG CATGCCCTTC ACCCTCCGC CCAATCTCAT GGGGCACTTC
360
AATTTTGGCA CGGTATCTG CAAGGCGGT TCTTACTCA TGGGGTGTG TGTAGTGTG
420
TTCAGGCTTA GCCTGTGGC CATGCCACTG GAGGATATA GGGCATCTG CCGACACTG
480
CAGGACGAG TGTGCGAAG GCGCTCCAC GCGCTGCGG TGAATTGAG CACGTGGCTG
540
CTGTCCGAG TACTCATGT GCCTTACCC GTGTACTG TGTGCAACC AGTGGGGCTT
600
CGTGTCTGC AGTGCATGA TGGCTGGCC AGTGCAGGG TCCGCCAGAC CTGTTCGTA
660
CTGTCTTTC TGCTTGTGTT CTTCATCCA GGTGTGTTA TGGCCGTGGC CTACGGGCTT
720
ATCTCTGGG AGCTTACTT AGGGCTTGC TTGTAGAGC AGATGACAG CGACAGCCAA
780
AGCAGGCTCC GAATACAGG CCGGCTGCA GGGGCTGTC ACCAGAACG GCGTTGCCG
840
CTGAGACTG GCGGGTTGG CAATGACAG GATGCTGCT AGTGCACACT TCCACGTTCC
900
CGGCTGCCC TGGAGCTGAC GAGGCTGAG GCTCTGGGC CAGGATCCG CTCGCCGCC

WO 00/22131

- 101 -

100 105 110
 Leu Pro Asn Leu Met Gly Thr Phe Ile Phe Gly Thr Val Ile Cys Lys
 115 120 125
 Ala Val Ser Tyr Leu Met Gly Val Ser Val Ser Val Ser Thr Leu Ser
 130 135 140
 Leu Val Ala Ile Ala Leu Glu Arg Tyr Ser Ala Ile Cys Arg Pro Leu
 145 150 155 160
 Gln Ala Arg Val Thr Gln Thr Arg Ser His Ala Ala Arg Val Ile Val
 165 170 175
 Ala Thr Trp Leu Leu Ser Gly Leu Leu Met Val Pro Tyr Pro Val Tyr
 180 185 190
 Thr Val Val Gln Pro Val Gly Pro Arg Val Leu Gln Cys Val His Arg
 195 200 205
 Trp Pro Ser Ala Arg Val Arg Gln Thr Trp Ser Val Leu Leu Leu
 210 215 220
 Leu Leu Phe Phe Ile Pro Gly Val Val Met Ala Val Ala Tyr Gly Leu
 225 230 235 240
 Ile Ser Arg Glu Leu Tyr Leu Gly Leu Arg Phe Asp Gly Asp Ser Asp
 245 250 255
 Ser Asp Ser Gln Ser Arg Val Arg Asn Gln Gly Gly Leu Pro Gly Ala
 260 265 270
 Val His Gln Asn Gly Arg Cys Arg Pro Glu Thr Gly Ala Val Gly Lys
 275 280 285
 Asp Ser Asp Gly Cys Tyr Val Gln Leu Pro Arg Ser Arg Pro Ala Leu
 290 295 300
 Glu Leu Thr Ala Leu Thr Ala Pro Gly Pro Gly Ser Gly Ser Arg Pro
 305 310 315 320
 Thr Gln Ala Lys Leu Leu Ala Lys Lys Arg Val Lys Arg Met Leu Leu
 325 330 335
 Val Ile Val Val Leu Phe Leu Cys Trp Leu Pro Val Tyr Ser Ala
 340 345 350
 Asn Thr Trp Arg Ala Phe Asp Gly Pro Gly Ala His Arg Ala Leu Ser
 355 360 365
 Val Ala Pro Ile Ser Phe Ile His Leu Leu Ser Tyr Ala Ser Ala Cys
 370 375 380
 Val Asn Pro Leu Val Tyr Cys Phe Met His Arg Arg Phe Arg Gln Ala
 385 390 395 400

WO 00/22131

- 100 -

960
 ACCGAGGCCA AGCTGTGGC TAAGAGCGC GTGACGNA TGTTGTGGT GATCGTTGTG
 1020
 CTTTTTTTC TGTTTGGTT GCCAGTTTAT AGTGCCACA CGTGGCGGC CTTTGATGCC
 5 1080
 CCGGTGCAC ACCGAGCACT CTCGGGTGT CCTATCTCTT TCATTCACTT GCTGAGCTAC
 1140
 GCCTCGGCT GTGTCAACCC CTTGGTCTAC TGCTTCATGC ACCGTGCTT TCGCCAGGCC
 1200
 TGCTGGAAA CTTGCGCTG CTGCTGCCCC CGGCTCCAC GAGCTGCCCC CAGGGCTCTT
 1260
 CCGATGAGG ACCCTCCAC TCCCTCCATT GCTTCGTGT CAGGCTTAG CTACACACC
 1320
 ATCAGCACAC TGGGCCCTGG CTGA
 15 1344
 (133) INFORMATION FOR SEQ ID NO:132:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 447 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant
 20
 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:
 Met Glu Leu Leu Lys Leu Asn Arg Ser Val Gln Gly Thr Gly Pro Gly
 1 5 10 15
 Pro Gly Ala Ser Leu Cys Arg Pro Gly Ala Pro Leu Leu Asn Ser Ser
 20 25 30
 Ser Val Gly Asn Leu Ser Cys Glu Pro Pro Arg Ile Arg Gly Ala Gly
 35 40 45
 Thr Arg Glu Leu Glu Leu Ala Ile Arg Ile Thr Leu Tyr Ala Val Ile
 50 55 60
 Phe Leu Met Ser Val Gly Gly Asn Met Leu Ile Ile Val Val Leu Gly
 65 70 75 80
 Leu Ser Arg Arg Leu Arg Thr Val Thr Asn Ala Phe Leu Leu Ser Leu
 85 90 95
 Ala Val Ser Asp Leu Leu Leu Ala Val Ala Cys Met Pro Phe Thr Leu

- 102 -

Cys Leu Glu Thr Cys Ala Arg Cys Cys Pro Arg Pro Arg Ala Arg
405 410 415

Pro Arg Ala Leu Pro Asp Glu Asp Pro Pro Thr Pro Ser Ile Ala Ser
420 425 430

Leu Ser Arg Leu Ser Tyr Thr Thr Ile Ser Thr Leu Gly Pro Gly
435 440 445

(134) INFORMATION FOR SEQ ID NO:133:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1014 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:133:

15 ATGAAACAGCA CATTATATGA AGAACAGCAT GACTGATC ACTATTGTT TCCATTGTT 60
TACATCTTG TGATTATAGT CAGCATCCA GCCATATATG GATCTGTG TGTCCTTTC 120
CTGCAAGCA AAGAGAAAG TGAATAGGA ATTACCTCT TCACTTGT ACTATCAGAT 180
TTACTATAG CATTACTCT CCTTATAG ATTGATATA CTGGAATATA AGACACTGG 240
ACTTCTCTC CTGCTGTG CAAGGAGT GCTTCTCA TGTACATGAA TTTTACAGC 300
20 AGACAGCAT TCCATCCTG CATTGCCGT GATCGGATTT TGCTGTGT CTACCTTTG 360
AAGTTTTT TCCTAGAGC AAGAGATTT GCATCATAG TCACTGTCT CATTGAGATA 420
TTGAAACCA TCTTCAATG TGTCAATGT TGGAAAGATG AAGAGTTGT TGATATTTG 480
GATGCCGAA AGCTAATTT TACTTATGC TATGACAAAT ACCCTTAGA GAATGGCAA 540
ATCAACTCA ACTTGTGAG GACGTGTACA GGTATAGCA TACCTTGT CACCATCTG 600
25 ATCTTAAC GGAAGCTTA CCAAGCTG CCGACATATA AAGCAGGGA AAGCAGGGA 660
AAGAGAGA TCAAAACT ACTTGTAGC ATCAGCTTA CTTTGCTT AGCTTACT 720
CCCTTCAAG TGATGTGCT GATTCGTGC ATTTAGAGC ATGCTGGA CTTCAGAGC 780
CAGCAGATT CTGGAAGCG AACTTACACA ATGTATGAA TCAAGTTG ATTACAAAGT 840
TTAAATGAG TTGCTATCC AATCTGTAC TGTTTGTTA CCGAAACAG AGATATGAT 900
30 ATGCGAATA TATTAATTT CTGACTGGG AGGTGTAATA CATCACAAG ACAAGAAAA 960
CGCATCTT CTGTGTAC AAGAGTACT ATGAGTTAG AGTCTCTGA GTCG 1014

- 103 -

(135) INFORMATION FOR SEQ ID NO:134:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 337 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(11) MOLECULE TYPE: protein

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:134:

10 Met Asn Ser Thr Cys Ile Glu Glu Gln His Asp Leu Asp His Tyr Leu 15
1 5 10 15

Phe Pro Ile Val Tyr Ile Phe Val Ile Ile Val Ser Ile Pro Ala Asn 20
20 25 30

Ile Gly Ser Leu Cys Val Ser Phe Leu Gln Ala Lys Lys Glu Ser Glu 35
35 40 45

Leu Gly Ile Tyr Leu Phe Ser Leu Ser Leu Ser Asp Leu Leu Tyr Ala 50
50 55 60

Leu Thr Leu Pro Leu Tyr Ile Asp Tyr Thr Trp Asn Lys Asp Asn Trp 65
65 70 75 80

Thr Phe Ser Pro Ala Leu Cys Lys Lys Gly Ser Ala Phe Leu Met Tyr Met 85
85 90 95

20 Asn Phe Tyr Ser Ser Thr Ala Phe Leu Thr Cys Ile Ala Val Asp Arg 100
100 105 110

Tyr Leu Ala Val Tyr Pro Leu Lys Phe Phe Phe Leu Arg Thr Arg 115
115 120 125

25 Arg Phe Ala Leu Met Val Ser Leu Ser Ile Tyr Ile Leu Glu Thr Ile 130
130 135 140

Phe Asn Ala Val Met Leu Trp Glu Asp Glu Thr Val Val Glu Tyr Cys 145
145 150 155 160

Asp Ala Glu Lys Ser Asn Phe Thr Leu Cys Tyr Asp Lys Tyr Pro Leu 165
165 170 175

30 Glu Lys Tyr Gln Ile Asn Leu Asn Leu Phe Arg Thr Cys Thr Gly Tyr 180
180 185 190

Ala Ile Pro Leu Val Thr Ile Leu Ile Cys Asn Arg Lys Val Tyr Gln 195
195 200 205

35 Ala Val Arg His Asn Lys Ala Thr Glu Asn Lys Glu Lys Lys Arg Ile 210
210 215 220

- 105 -

ATTGATTAATG TCATGTGACTC GGTGATCTGT AGCTCCTTGC TTGCATCCAT TTGCAGCCTG
420
CTTCAATATG CAGTGGACAG GTACTTTACT ATCTTCTATG CTCCTCAGTA CCATAACATT
480
5 ATGACAGTTA AGCGGGTTGG GATCAGCATA AGTTGTATCT GGGCAGCTTG CAGGGTTTCA
540
GGCATTTTGT TCATCAATTA CTCAGATAGT AGTGTGTGTA TCATGTGCTT CATACCATG
600
TTCTTCACCA TGCTGGCTCT CATGGCTTCT CTCTATGTCC ACATGTTTCT GATGSCCAGG
10 660
CTTCACATTA AGAGGATTCG TGTCTCCTCC GGCATGTGCT CCATCCGCCA AGGTGCCAAT
720
ATGAGGAGAA AATTTACCTT GACCATCTCG ATTGGCTGCT TTGTTGTCTG CTGGGCCCCA
780
15 TTCTTCTCC ACTTATATT CTACATCTCT TGTCTCTCAGA ATCCATATTG TGTGTGCTTC
840
ATGCTCACT TTAATCTGTA TCTCACTAGT ATCAATGTA ATTCATCAT CGATCCTCTG
900
ATTATGAC TCCGGAGTCA AGAATGAGG AATACCTTCA AAGATCAT CTGTTGCTAT
20 960
CCCTTGGGAG GCCTTTGGA CTGTCTAGC AGATATTAA
999

(137) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 332 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

Met Val Asn Ser Thr His Arg Gly Met His Thr Ser Leu His Leu Trp
1 5 10 15
Asn Arg Ser Ser Tyr Arg Leu His Ser Asn Ala Ser Glu Ser Leu Gly
20 25 30
Lys Gly Tyr Ser Asp Gly Cys Tyr Glu Gln Leu Phe Val Ser Pro
35 40 45

- 104 -

Lys Lys Leu Leu Val Ser Ile Thr Val Thr Phe Val Leu Cys Phe Thr
225 230 235 240
Pro Phe His Val Met Leu Ile Arg Cys Ile Leu Glu His Ala Val
245 250 255
5 Asn Phe Glu Asp His Ser Asn Ser Gly Lys Arg Thr Tyr Thr Met Tyr
260 265 270
Arg Ile Thr Val Ala Leu Thr Ser Leu Asn Cys Val Ala Asp Pro Ile
275 280 285
Leu Tyr Cys Phe Val Thr Glu Thr Gly Arg Tyr Asp Met Trp Asn Ile
290 295 300
Leu Lys Phe Cys Thr Gly Arg Cys Asn Thr Ser Gln Arg Gln Arg Lys
305 310 315 320
Arg Ile Leu Ser Val Ser Thr Lys Asp Thr Met Glu Leu Glu Val Leu
325 330 335
15 Glu
340

(136) INFORMATION FOR SEQ ID NO:135:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 999 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

ATGGTGAAT CCACCCACCG TGGATGACAC ACTTCTCTGC ACTCTGGAA CCGAGCAGT
60
TACAGACTGC ACAGCAATGC CAGTGAGTCC CTGGAAGAG GCTACTCTGA TGGAGGGTGC
120
TACGAGCAAC TTTTGTCTC TCCTGAGGTG TTTGTGACTC TGGGTGTCAAT CAGCTTGTG
30 180
GAGATATCT TAGTGATTGT GGCAATAGCC AAGAACAAGA ATCTGCATTC ACCCATGTAC
240
TTTTTCATCT GCAGCTTGGC TGTGGCTGAT ATGCTGGTGA GCGTTTCAAA TGGATCAGAA
300
35 ACCATTATCA TCACCTTATT AAACAGTACA GATACGGATG CACAGAGTTT CACAGTGAAT
360

- 106 -

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

(138) INFORMATION FOR SEQ ID NO:137:

- 107 -

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:137:
GCCAATATGA AGGAAAT TACTGACC ATC
33

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:138:
CTCCTCGGT CCTCTATG TTGTCAGAG T
31

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1842 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:139:
ATGGGGCCCA CCTAGGCGT TCCACCCCC TATGCTGTGA TTGCTGTGA GCTACCCAG
60
CCAGATPAC CACGCGCTT ATCATCTT ATGTTGCG CGATGTAT CACATCGT
120
GTAACTTAA TCGCACTC CATGTCAT TTGCTGTGA CGAAGAAC GAAGCTCGG
180
AATCTGGA ACATCTTGT GTCAGTTC TCTGTGCG ATATGCTGT GGCATCTAC
240
CCATACCTT TGATCTGCA TCCATGTC ATGCGGCGT GGATCTGAG CCAATACAG
300
TCCAGATG TCGGCTCAT CACAGGCGT AGTGTGTG GCTCATCTT CAACATGCTG
360

- 108 -

GCAATCGCTA TCAACCGTTA CTGTACATC TCCACAGCC TCCAGTAGA ACGGATCTTC 420
AGTGTGGCCA ATACTGCGAT CTACTGTGC ATCAGCTGGA TCAATGACGT CCGTGGCTGTC 480
CTGGCCAACA TGTACATTTG CACCATCGAG TAGATCCTC GCACCTACAC CTGCATCTTC 540
AACTATCTGA ACAACCTGT CTTCATGTT ACCATCGTCT GCATCCACTT CGTCTCTCCT 600
CTCTCTCATCG TGGTTTCTG CTACTGAGG ATCTGAGCA AGTGTCTGGC GCGCCGTGAC 660
CTGCGAGGCG AGAATCTCTGA CAACCACTT GCTGAGGTTT GCAATTTTCT AACCATGTTT 720
GTGATCTTCC TCTCTTTTC AGTGTCTG TGCCCTATCA ACGTCTCTAC TGTCTTGTG 780
GCTGTCAATC CGAAGAGAT GCGAGGCAAG ATCCCAACT GCGTTATCT TGCAGCTTAC 840
TTATAGCCT ACTTCAACAG CTGCTCTAAC GCTGTGATCT ACGGCTCTCT CAATGAGAT 900
TTCCGAGAG ATACTGAGC CATCTTCCAT GCTATGCGC ACCATATCAT ATTCTTCCCT 960
GGCTCTATCA GTGATATTCG TGAGATGCG GAGGCCCGTA CCTGGCCCG CCGCCGTGCC 1020
CATGCTGCG ACCAGCTCG TGACAGAC CGTGCCGATG CCGTCTCTGC TGTGGAGAA 1080
ACCCGATGA ATGTCCGAA TGTTCATTA CTTGTGATG CTGCACTGG CCACCCGAC 1140
CGTGCCTCTG GCCACCTTAA GCCCAATTC AGATCTCTCT CTGCTATTCG CAATCTGCC 1200
TTCTACCCAC ACAGTCTGT CTTTAGCCAC TCCAGGCTG CCGTCTGTCA CCGTACGCT 1260
GTCTCTGCC ACTCCAAGC TGCTCTGCT CACCCCAAGT CTGCCACTGT CTACCTTAAG 1320
CCTGCTCTG TCCATTTCAA GGTGACTCT GTCCATTTCA AGGTGACTC TGTCCATTTT 1380
AAGCTGACT CTGTTCAATTT CAAGCTGCT TCCAGCAACC CCAAGCCCAT CACTGGCCAC 1440
CATGCTCTG CTGGCAGCA CTCCAAGT GCTTCCAGTG CTGCCACCAG CCACCTTAA 1500
CCCATCAAGC CAGCTACCAAG CCATGCTGAG CCCACCACTG CTGACTATCC CAAGCTGCC 1560
ACTACCAACC ACCCTAAGCC CGTGTCTGT GACAACCTTG AGCTCTCTGC CTCCCATGTC 1620
CCGAGATCC CTGCCATTGC CCACCTGTG TCTGACGACA GTGACTCTCC TGAGTGGCC 1680
TCTAGCCCTG CCGCTGGCC CACCAAGCT GCTGCCAGCC AGCTGAGTGC TGACACCATC 1740
GCTGACCTTC CTGACCTTAC TGTAGTCACT ACCAGTACCA ATGATTAACCA TGATGCTGTG 1800
GTGTTGTATG TTGAAGATGA TCTGTANGAA ATGGTGTGT GA 1842

(141) INFORMATION FOR SEQ ID NO:140:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 613 amino acids

(B) TYPE: amino acid

- 109 -

(C) STRANDEDNESS:
(D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys 15
1 5 10
Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe 30
20 25 30
Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met 45
35 40 45
Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn 60
50 55 60
Ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr 75
65 70 75 80
Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Gly Trp Asp Leu 95
85 90 95
Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val 110
100 105 110
Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys 125
115 120 125
Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn 140
130 135 140
Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val 160
145 150 155 160
Leu Pro Asn Met Tyr Ile Gly Thr Ile Glu Tyr Asp Pro Arg Thr Tyr 175
165 170 175
Thr Cys Ile Phe Asn Tyr Leu Asn Asn Pro Val Phe Thr Val Thr Ile 190
180 185 190
Val Cys Ile His Phe Val Leu Pro Leu Leu Ile Val Gly Phe Cys Tyr 205
195 200 205
Val Arg Ile Trp Thr Lys Val Leu Ala Ala Arg Asp Pro Ala Gly Gln 220
210 215 220 225
Asn Pro Asp Asn Gln Leu Ala Glu Val Arg Asn Phe Leu Thr Met Phe 240
225 230 235 240
Val Ile Phe Leu Leu Phe Ala Val Cys Trp Cys Pro Ile Asn Val Leu 255
245 250 255

- 110 -

Thr Val Leu Val Ala Val Ser Pro Lys Glu Met Ala Gly Lys Ile Pro
260 265 270

Asn Tyr Leu Tyr Leu Ala Ala Tyr Phe Ile Ala Tyr Phe Asn Ser Cys
275 280 285

Leu Asn Ala Val Ile Tyr Gly Leu Leu Asn Glu Asn Phe Arg Arg Glu
290 295 300

Tyr Tyr Thr Ile Phe His Ala Met Arg Arg His Pro Ile Ile Phe Phe Pro
305 310 315 320

Gly Leu Ile Ser Asp Ile Arg Glu Met Glu Glu Ala Arg Thr Leu Ala
325 330 335

Arg Ala Arg Ala His Ala Arg Asp Glu Ala Arg Glu Glu Asp Arg Ala
340 345 350

His Ala Cys Pro Ala Val Glu Glu Thr Pro Met Asn Val Arg Asn Val
355 360 365

Pro Leu Pro Gly Asp Ala Ala Ala Gly His Pro Asp Arg Ala Ser Gly
370 375 380

His Pro Lys Pro His Ser Arg Ser Ser Ala Tyr Arg Lys Ser Ala
385 390 395 400

Ser Thr His His Lys Ser Val Phe Ser His Ser Lys Ala Ala Ser Gly
405 410 415

His Leu Lys Pro Val Ser Gly His Ser Lys Pro Ala Ser Gly His Pro
420 425 430

Lys Ser Ala Thr Val Tyr Pro Lys Pro Ala Ser Val His Phe Lys Gly
435 440 445

Asp Ser Val His Phe Lys Gly Asp Ser Val His Phe Lys Pro Asp Ser
450 455 460

Val His Phe Lys Pro Ala Ser Ser Asn Pro Lys Pro Ile Thr Gly His
465 470 475 480

His Val Ser Ala Gly Ser His Ser Lys Ser Ala Phe Ser Ala Ala Thr
485 490 495

Ser His Pro Lys Pro Ile Lys Pro Ala Thr Ser His Ala Glu Pro Thr
500 505 510

Thr Ala Asp Tyr Pro Lys Pro Ala Thr Thr Ser His Pro Lys Pro Ala
515 520 525

Ala Ala Asp Asn Pro Glu Leu Ser Ala Ser His Cys Pro Glu Ile Pro
530 535 540

Ala Ile Ala His Pro Val Ser Asp Ser Asp Leu Pro Glu Ser Ala

- 111 -

545 550 555 560

Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Glu Leu Glu
565 570 575

Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser
580 585 590

Thr Asn Asp Tyr His Asp Val Val Val Asp Val Glu Asp Asp Pro
595 600 605

Asp Glu Met Ala Val
610

(142) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1842 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:141:

ATGGGGGCCA CCTAGCGGT TCCACCCCC TAGGCTGTA TTGGCTGTA GTCACCCG 60

CCAGATTACC CACCGGCTCT AATCATCTTT ATGTTCTGCG CGATGGTAT CACATCTT 120

GTAGACCTAA TCGGCAACTC CAGTGTCAAT TTGGCTGTA CGAAGAACAA GAAAGTCCGG 180

AATCTGGCA ACATCTTGGT GGTCACTCTC TCTGTGGCG ATATGCTGGT GGCATCTTAC 240

CCATACCTTT TGATGCTGCA TGCCATGTC ATTGGGGGCT GGGATCTGAG CAGTTACAG 300

TGCCAGATGG TCGGGTTCAAT CACAGGGCTG AGTGGGTGCG GCTCCATCTT CAACATCGTG 360

GCAATGCTTA TCACCGCTTA CTGTCTATATC TGGCAGAGCC TCCAGTAGCA ACGGATCTTC 420

AGTGTAGGCA ATACCTGCAAT CTACTGTGTC ATCAGCTGGA TCATGACCGT CTTGGCTGTC 480

CTGCCAACAA TGTACATGAG CACCATGAG TAGATCTTC GCACTTACAC CTGCATCTTC 540

AACATCTGA ACAACCTGT CTTCACGTGT ACCATGTGCT GCATCACTT GTCTCTCCCT 600

CTCCATGAG TGGGTTTCTG CTACGTGAGG ATCTGAGCA AAGTGTGAGC GAGCCGTGAC 660

CTTGAGAGGC AGAATCTGTA CAACCACTT GCTGAGGTTT GCATTAACCT AACCATGTTT 720

GTGATCTTCC TCTCTTTGCG AGTGTGCTGG TGCCCTATCA ACGTGTACAC TGTCTTGGTG 780

GCTGTCACTC CGAAGAGAT GGCAGGAGAG ATCCCACTACT GCGTTATCTA TGCAGCTTAC 840

TTCTATGACCT ACTTCAACAG CTGCTCAAC GCTGTGATCT AGGGCTCTCT CAATGAGAAT 900
 TTCCGAAGAG AATACTGGAC CATCTTCCAT GCTATGGGC ACCCTATCAT ATTCTTCTCT 960
 GGCCTCATCA GTGATATTTC TGAGATGCAG GAGGCCCGTA CCTTGGCCCG CGCCCTGGCC 1020
 CATGCTCGCG ACCAAGCTCG TGACAAAGAC CGTGCCCATG CCTGTCTGTC TGTGGAGGAA 1080
 5 ACCCCGATGA ATGTCCGGBA TGTTCATTA CTGTGGTAGT CTGAGCTGG CCACCCCGAC 1140
 CGTGCCTCTG GCCACCTTAA GCCCCATTCC AGATCTCTCT CTGCCTATCG CAATCTGCC 1200
 TCTACCCAGC ACAAGTCTGT CTTTACGAC TCCAAAGGCTG CCTCTGCTCA CCTCAAGCCT 1260
 GTCTCTGGCC ACTCAGGCC TGCTCTGTGT CAGCCCAAGT CTGCCACTGT CTACCCCTAAG 1320
 CCTGCCTCTG TCCATTTCAA GGTGACTCT GTCCATTTC AAGGTGACTC TGTCCATTTC 1380
 10 AAGCCTGACT CTGTTCATTT CAAGCCTGCT TCCAGCAACC CCAAGCCCAT CACTGSCCAG 1440
 CATGTCTCTG CTGGCAGCCA CTCGAAGTCT GCCTTCAATG CTGCCACCAAG CCACCTTAA 1500
 CCCATCAGC CAGCTACCAAG CCATGCTGAG CCCACCACTG CTGACTATTC CAAGCCTGCC 1560
 ACTTACCAGC ACCTTAAGCC CGCTGTCTGT GACAAACCCTG AGCTCTCTGC CTCCCATTCG 1620
 CCGGAGATCC CTGCCATTCG CCACCTCTG TCTGACGACA GTGACCTCCC TGAGTCGGCC 1680
 15 TCTAGCCCTG CCGCTGGCC CACCAAGCT GCTGCCAGCC AGCTGGAGTC TGACACCATC 1740
 GCTGACCTTC CTGACCTTAC TGTAGTCACT ACCAGTACCA ATGATTACCA TGTGTGCTG 1800
 GTTGTGTATG TTGAAGATGA TCTGTATGAA ATGGCTGTGT GA 1842

(143) INFORMATION FOR SEQ ID NO:142:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 613 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: not relevant

(ii) MOLECULE TYPE: protein

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Met Gly Pro Thr Leu Ala Val Pro Thr Pro Tyr Gly Cys Ile Gly Cys
 1 5 10 15
 Lys Leu Pro Gln Pro Glu Tyr Pro Pro Ala Leu Ile Ile Phe Met Phe
 20 25 30
 Cys Ala Met Val Ile Thr Ile Val Val Asp Leu Ile Gly Asn Ser Met
 35 40 45

Val Ile Leu Ala Val Thr Lys Asn Lys Lys Leu Arg Asn Ser Gly Asn
 50 55 60
 ile Phe Val Val Ser Leu Ser Val Ala Asp Met Leu Val Ala Ile Tyr
 65 70 75 80
 5 Pro Tyr Pro Leu Met Leu His Ala Met Ser Ile Gly Trp Asp Leu
 85 90 95
 Ser Gln Leu Gln Cys Gln Met Val Gly Phe Ile Thr Gly Leu Ser Val
 100 105 110
 Val Gly Ser Ile Phe Asn Ile Val Ala Ile Ala Ile Asn Arg Tyr Cys
 115 120 125
 Tyr Ile Cys His Ser Leu Gln Tyr Glu Arg Ile Phe Ser Val Arg Asn
 130 135 140
 Thr Cys Ile Tyr Leu Val Ile Thr Trp Ile Met Thr Val Leu Ala Val
 145 150 155 160
 15 Leu Pro Asn Met Tyr Ile Gly Thr Ile Glu Tyr Asp Pro Arg Thr Tyr
 165 170 175
 Thr Cys Ile Phe Asn Tyr Leu Asn Asn Pro Val Phe Thr Val Thr Ile
 180 185 190
 Val Cys Ile His Phe Val Leu Pro Leu Ile Val Gly Phe Cys Tyr
 195 200 205
 Val Arg Ile Trp Thr Lys Val Leu Ala Ala Arg Asp Pro Ala Gly Gln
 210 215 220
 Asn Pro Asp Asn Gln Leu Ala Glu Val Arg Asn Lys Leu Thr Met Phe
 225 230 235 240
 25 Val Ile Phe Leu Leu Phe Ala Val Cys Trp Cys Pro Ile Asn Val Leu
 245 250 255
 Thr Val Leu Val Ala Val Ser Pro Lys Glu Met Ala Gly Lys Ile Pro
 260 265 270
 30 Asn Trp Leu Tyr Leu Ala Ala Tyr Phe Ile Ala Tyr Phe Asn Ser Cys
 275 280 285
 Leu Asn Ala Val Ile Tyr Gly Leu Leu Asn Glu Asn Phe Arg Arg Glu
 290 295 300
 Tyr Trp Thr Ile Phe His Ala Met Arg His Pro Ile Ile Phe Phe Ser
 305 310 315 320
 35 Gly Leu Ile Ser Asp Ile Arg Glu Met Gln Glu Ala Arg Thr Leu Ala
 325 330 335
 Arg Ala Arg Ala His Ala Arg Asp Gln Ala Arg Glu Gln Asp Arg Ala

- 114 -

340 345 350
His Ala Cys Pro Ala Val Glu Thr Pro Met Asn Val Arg Asn Val
355 360 365
Pro Leu Pro Gly Asp Ala Ala Gly His Pro Asp Arg Ala Ser Gly
370 375 380
His Pro Lys Pro His Ser Arg Ser Ser Ala Tyr Arg Lys Ser Ala
385 390 395 400
Ser Thr His His Lys Ser Val Phe Ser His Ser Lys Ala Ala Ser Gly
405 410 415
His Leu Lys Pro Val Ser Gly His Ser Lys Pro Ala Ser Gly His Pro
420 425 430 435
Lys Ser Ala Thr Val Tyr Pro Lys Pro Ala Ser Val His Phe Lys Ala
435 440 445
Asp Ser Val His Phe Lys Gly Asp Ser Val His Phe Lys Pro Asp Ser
450 455 460
Val His Phe Lys Pro Ala Ser Ser Asn Pro Lys Pro Ile Thr Gly His
465 470 475 480
His Val Ser Ala Gly Ser His Ser Lys Ser Ala Phe Asn Ala Thr
485 490 495
Ser His Pro Lys Pro Ile Lys Pro Ala Thr Ser His Ala Glu Pro Thr
500 505 510
Thr Ala Asp Tyr Pro Lys Pro Ala Thr Thr Ser His Pro Lys Pro Ala
515 520 525
Ala Ala Asp Asn Pro Glu Leu Ser Ala Ser His Cys Pro Glu Ile Pro
530 535 540
Ala Ile Ala His Pro Val Ser Asp Asp Ser Asp Leu Pro Glu Ser Ala
545 550 555 560
Ser Ser Pro Ala Ala Gly Pro Thr Lys Pro Ala Ala Ser Glu Leu Glu
565 570 575
Ser Asp Thr Ile Ala Asp Leu Pro Asp Pro Thr Val Val Thr Thr Ser
580 585 590
Thr Asn Asp Tyr His Asp Val Val Val Asp Val Glu Asp Asp Pro
595 600 605
Asp Glu Met Ala Val
610

(144) INFORMATION FOR SEQ ID NO:143:

- 115 -

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA (genomic)
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:143:
GCTGAGGTTT GCATTAACCT AACCATGTTT GTG
(145) INFORMATION FOR SEQ ID NO:144:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA (genomic)
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:144:
CTCTCTCGGT CCTCTATCG TTGTCAAGAG T
(146) INFORMATION FOR SEQ ID NO:145:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA (genomic)
(1v) ANTI-SENSE: NO
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:145:
TTAGATATCG GAGCCACCC TAGCGGT
(147) INFORMATION FOR SEQ ID NO:146:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA (genomic)

-116 -

(iv) ANTI-SENSE: YES

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:146:

GGTACCCCA CAGCCATTC ATCAGGAC